Improving real-time MRI for the clinical assessment of velar closure and velopharyngeal motion during speech

Andreia Calisto de Freitas

Submitted in partial fulfilment of the requirements of the Degree

of

Doctor of Philosophy

QUEEN MARY UNIVERSITY OF LONDON

School of Medicine and Dentistry

William Harvey Research Institute

London, United Kingdom

May 2018

Statement of originality

I, Andreia Calisto de Freitas, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third partys copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date:

Details of collaboration and publications: open-source software made available by the authors was used to partly develop the work presented in Chapters 5 and 6. Full acknowledgement is given. Colleague Matthieu Ruthven performed

some visual image quality scoring as a second observer for the experimental work described in Chapter 4. Full list of publications is given in Section 1.2.

Abstract

Magnetic Resonance Imaging (MRI) has been used to provide high-resolution tomographic information, valuable in the study of static vocal tract. However, speech does not present a static behaviour but relies on the continuous and dynamic interaction of the vocal tract articulators and neighbouring tissues. Thus, this could make real-time MRI (rt-MRI) an essential tool to assess speech, with numerous advantages over the current clinical techniques. However, using rt-MRI to image the upper vocal tract remains challenging; the motion of the articulators, including the velum is fast while MRI data acquisition is slow thus inherently limiting temporal resolution. Additionally, an intrinsic loss in SNR, spatial resolution and/or visual image quality is present. The main focus of this thesis is to increase clinical reliability of rt-MRI in speech by investigating novel methodologies for the imaging of velopharyngeal motion. Firstly, commercial rt-MRI protocols at 1.5 T and 3 T were compared, regarding image quality and temporal resolution compromise. Optimal imaging protocols were suggested for both field strengths. This provided a starting point for future clinical translation and the use of commercial and currently available protocols to image velopharyngeal motion. Furthermore, imaging of velopharyngeal motion was further improved with non-standard acquisition methods, such as non-Cartesian sampling and more advanced reconstruction schemes. An improved imaging protocol for the assessment of velopharyngeal motion was suggested. This was based on a highly accelerated radial trajectory with a novel parallel imaging reconstruction method (radial tt-GRAPPA). The suggested protocol not only allowed for improved image quality and image sharpness,

but it was also viable for future clinical translation regarding offline computation times compared to other reconstruction methodologies also investigated in this thesis. In summary, this thesis added some novel insights into the field of speech rt-MRI, presenting improved and time efficient imaging protocols, adequate for the assessment of velopharyngeal motion.

Acknowledgments

Four years have passed and what a journey it has been!

I would like to start by saying that the work present in this thesis would not have been possible without the contribution of Dr. Marc Miquel and Dr. Redha Boubertakh. I am forever grateful for all the guidance, scientific knowledge and support they have demonstrated throughout this journey. I am also grateful to CLEFT for funding the PhD project and making it possible for me to dedicate 3 years to the field of speech MRI.

A special Obrigado (thank you) to my dearest Portuguese friend Zé, for always making me laugh and sharing the funniest moments together. Thank you to Matthieu for all the time spent in the scanner together and all the interesting discussions shared on the speech project. Many thanks to everyone for the fun afternoons playing table football, as it made me realise that I might be a much better scientist than footballer!

Also many thanks to Professor Steffen Petersen and the entire team at CVBRU. Thank you to everyone who took the time to volunteer for me, you made this work possible.

Thank you to my dearest and always in my heart childhood friends, Ritinha, Gil, Santos e Edu. Thank you for always being there for me even when it meant being 1500 km away for these past few years and absent in so many important life moments. Hopefully we can get together soon to celebrate this important mark in my life!

I am also very thankful to all my fellow Londoners. Thank you for all the

words of encouragement, moments of fun and pints shared together. A special thanks to Catarina, Hugo, Juliane, Rui, Johny, Carlinha and Andreia.

It goes without saying, how much in my heart I appreciate the unconditional love and support Luis has showed me throughout these years. My name might be solely written on the front page of this thesis but he was the foundation I needed to complete this journey. It will take a lifetime to repay you.

Lastly but definitely not least, thank you mum, dad and little sis (although taller than me!). My strength, my roots and my love. Thank you for always believing in me even when I didn't believe it myself.

Dedicado aos meus avós Laura, Manuel, Tina e Feliciano.

"I am among those who think
that science holds great beauty."

Marie Curie

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Magnetic Resonance Imaging (MRI) is a versatile and powerful imaging modality, that provides non-invasive, free of ionising radiation and flexible selection of tissue contrast tomographic information.

A number of disorders, both acquired and congenital, may affect the vocal tract articulators and hamper phonation of natural speech. These may include but not be limited to: vocal tract cancer, dysphagia, neurological disorders and orofacial clefts. Orofacial clefts are a set of congenital deformities that may encompass disorders of the lips, velum, pharynx and/or palate; affecting about 9.92 per 10,000 births worldwide [1]. Velopharyngeal insufficiency (also known as VPI), common in orofacial clefts, is a speech impairment resulting from the incomplete closure of the velopharyngeal port. Clinical assessment of VPI relies on a multi-disciplinary approach, which may include the speech therapists perceptual evaluation and imaging techniques (most commonly x-ray videofluoroscopy and/or nasendoscopy).

Although x-ray videofluoroscopy offers high spatial-temporal resolution; soft tissue contrast, such as the one of the velum, is relatively poor. Another major disadvantage concerns the repetitive exposure to ionising radiation. Additionally, the performance of nasendoscopy with a nasal scope is a rather invasive procedure, particularly for younger patients, and only qualitative measurements of the articulators is possible [2]. Limitations of these techniques have strongly fomented

the interest in using MRI for the investigation of speech and the upper vocal tract, as summarised in a review of the field [3]. MRI could potentially provide tomographic images with improved soft issue contrast (ideal for vocal tract imaging) with a flexible selection of imaging planes without the need to reposition the patient.

Furthermore, speech does not present a static behaviour but relies on moving vocal articulators that continuously interact with each other and neighbouring tissues. Thus, static MRI of the upper vocal tract, although valuable and interesting [3], can only provide limited information on the production of speech and velopharyngeal motion. An increased number of studies have used real-time MRI (rt-MRI) to image the upper airway during speaking, singing and swallowing, both in asymptomatic [4–8] and VPI individuals [9–11].

Dynamic imaging of the upper vocal tract during speech with rt-MRI is still an open field of research, holding great potential for future clinical translation in the assessment of speech disorders. However, many constraints still need to be addressed and further explored before clinical implementation is viable. Velopharyngeal motion is characterised by the rapid transition of the velum between rest and elevated position during velopharyngeal closure. Furthermore, dynamic imaging of the upper vocal tract is exceptionally challenging due to the presence of large areas of air-tissue boundary, thus being particularly prone to off-resonance effects.

In this thesis, I set out to explore some of the aspects currently hampering the use of rt-MRI in speech assessment. The main motivation for this work is the need for reliable acquisition and reconstruction methodologies to provide adequate spatial-temporal imaging to the assessment of velopharyngeal motion in speech and consequent future clinical translation.

Firstly, this thesis investigates the use of commercially available MRI tools, such as standard commercially-available sequences and reconstruction algorithms, to suggested optimal rt-MRI protocols to dynamically image velopharyngeal mo-

tion. This could provide an easy to implement and reproducible protocol across clinical centres to aid in future clinical translation. In addition, it may also provide some insight and a starting point to researchers and physicians interested in using dynamic MRI for the study of speech. A variety of commercially available protocols were compared and suggested for the two most commonly used clinical MR scanners, 1.5 T and 3 T field strength. Experimental work referring to this will be presented and discussed in Chapter 4.

After commercially available and easily reproducible imaging protocols have been compared and proposed for speech assessment, I set out to answer the following question: "How can we further improve rt-MRI of speech?". For this, novel acquisition and reconstruction methodologies will then be explored in order to improve rt-MRI of speech. Due to the efficient coverage of k-space and resilience to higher degrees of under-sampling, non-Cartesian imaging was the focus on all following experimental work.

Rt-MRI data of speech was acquired with moderately accelerated radial trajectory (also with an implementation of a golden-angle scheme) and image data was reconstructed offline with a novel iterative SENSE-based method with TGV regularisation. Although results obtained with TGV-SENSE iterative reconstruction and radial golden-angle could potentially be of interest for research speech studies, direct future clinical translation was still prohibitive due to the time required to reconstruct and process image data offline in current set-ups. Thus, I hypothesised that a GRAPPA based method could greatly decrease the necessary computation time while still providing adequate spatial-temporal resolution rt-MRI data sets of speech. Thus, the last experimental work in this thesis refers to the improved 2D rt-MRI of speech using accelerated radial data reconstructed with a novel reconstruction method known as through-time GRAPPA (tt-GRAPPA).

At the time that this experimental work was undertaken, no other description of this method applied to the dynamic imaging of speech was present in the current literature. Original descriptions of tt-GRAPPA were firstly introduced by

Seiberlich et al. [12] in 2011 to perform dynamic imaging of the heart. Although some challenges are particular to imaging the vocal tract, I hypothesised that the tt-GRAPPA method previously described for dynamic heart imaging could also bring improvement to the dynamic imaging of speech. Thus, an accelerated radial acquisition reconstructed with a tt-GRAPPA and hybrid GRAPPA scheme was investigated.

1.1 Thesis outline

This dissertation explores the use of rt-MRI for the assessment of velopharyngeal motion and investigates novel methodologies to improve spatial-temporal depiction. The content of the remaining chapters are outlined below:

Chapter 2: MRI physics

The main principles of MR physics, pulse sequences, k-space sampling and reconstruction relevant to understand the experimental work undertaken are presented in this chapter.

Chapter 3: Speech and Imaging the Upper Vocal Tract

This chapter presents a brief review on the clinical subject at focus in this thesis, most specifically human speech and the upper vocal tract. A description of the current clinical assessment of speech and translation to the use of MRI is also given.

Chapter 4: Visualising and Assessing Velopharyngeal Motion with Commercial Cartesian and non-Cartesian rt-MRI

This experimental chapter focuses on investigating rt-MRI protocols for the assessment of velopharyngeal motion and closure. Such protocols are optimised and compared using at this point only commercially available sequences, hardware and reconstruction algorithms, to ensure direct translation to clinical practice.

Recommended protocols are suggested for imaging at 1.5 T and 3 T.

Chapter 5: Feasibility of Accelerated Golden-Angle Radial Imaging with TGV-regularised Reconstruction in Velopharyngeal Imaging

This chapter demonstrates the application of TGV-regularised iterative reconstruction of accelerated radial speech rt-MRI.

Chapter 6: Improved 2D rt-MRI of Speech using Accelerated Radial tt-GRAPPA in Velopharyngeal Imaging

In this experimental chapter, I set out to investigate the use of accelerated non-Cartesian imaging and parallel imaging reconstruction to improve spatial-temporal depiction of velopharyngeal motion. The proposed methodology relies on a highly accelerated radial sampling scheme and a through-time GRAPPA reconstruction. Its application to dynamic imaging of speech is demonstrated.

Chapter 7: Conclusions and Future Work

The final chapter summarises the conclusions and contributions of the experimental works undertaken in this thesis. Possible future work is also discussed.

1.2 List of publications

The experimental work undertaken in this thesis has been partially published in peer-reviewed journal articles and presented at international conferences in the field:

Peer-reviewed journal articles

A C Freitas, M Ruthven, R Boubertakh, M E Miquel. Real-time speech MRI: commercial Cartesian and non-Cartesian sequences at 3T and feasibility of offline TGV reconstruction to visualise velopharyngeal motion. *European Journal of Medical Physics*, 2018. (*in press*)

A C Freitas, M Wylezinska, M J Birch, S E Petersen, M E Miquel. Comparison of Cartesian and non-Cartesian real-time MRI sequences at 1.5T to assess velar motion and velopharyngeal closure during speech. *PLoS ONE* 11(4), 2016.

M E Miquel, **A C Freitas**, M Wylezinska. Evaluating velopharyngeal closure with real-time MRI. *Pediatric Radiology.* 45: 941-942, 2014.

Conference Proceedings

A C Freitas, M Ruthven, R Boubertakh, M E Miquel. Improved real-time MRI to visualise velopharyngeal motion during speech using accelerated radial through-time GRAPPA. *Magnetic Resonance Materials in Physics Biology and Medicine*, ESMRMB Barcelona; 2017, 30(Suppl 1): S17-S18. (selected for oral presentation)

M Ruthven, **A C Freitas**, R Boubertakh, M E Miquel. Velum movement in speech: can real-time MRI provide new insights? *Magnetic Resonance Materials in Physics Biology and Medicine*, ESMRMB Barcelona; 2017, 30(Suppl 1): S21-S22. (selected for oral presentation)

A C Freitas, M Wylezinska, R Boubertakh, M E Miquel. Cartesian and non-Cartesian 3T real-time speech MRI: a comparison of image quality in the assessment of velopharyngeal closure. *Magnetic Resonance Materials in Physics Biology and Medicine*, ESMRMB Vienna; 2016, 29(Suppl 1): 158 (selected for oral presentation)

M Ruthven, **A C Freitas**, S F Keevil, M E Miquel. Real-time speech MRI: what is the optimal temporal resolution for clinical velopharyngeal closure assessment? *Proceedings of 24th Annual meeting of International Society for Magnetic Resonance in Medicine*, Singapore; 2016, 24: 3208.(poster)

A C Freitas, M Ruthven, M Wylezinska, R Boubertakh, M E Miquel. Cartesian

and non-Cartesian real-time MRI for the assessment of velopharyngeal closure during speech. in *8th Annual Workshop on Biomedical Engineering (WBME2016)*, Lisbon, Portugal; 2016. (prize for best abstract)

A C Freitas, M Wylezinska, M J Birch, S E Petersen, M E Miquel. Real-time speech MRI: a comparison of Cartesian and non-Cartesian sequences. *Proceedings of 23rd Annual meeting of International Society for Magnetic Resonance in Medicine*, Toronto; 2015, 23: 655. (selected for oral presentation)

M Wylezinska, **A C Freitas**, M J Birch, M E Miquel. k-t BLAST/k-t FOCUSS in real-time imaging of the soft palate during speech. *Proceedings of 23rd Annual meeting of International Society for Magnetic Resonance in Medicine*, Toronto; 2015, 23: 2302. (poster)

Chapter 2

MRI background

The main focus of this experimental thesis is to use the advantages of MRI as an imaging modality to image and assess velopharyngeal motion during natural speech. Therefore, an understanding of the MRI physics is essential before further improvements can be proposed to the field of speech MRI. The present chapter introduces the underlying theory of MR physics as well as image acquisition and reconstruction concepts necessary to support the subsequent experimental work.

2.1 Principles of nuclear magnetic resonance

MR imaging exploits a quantum mechanics property of subatomic particles known as the nuclear spin. Images can be produced from many nuclei, such as the sodium-23 (^{23}Na), phosphorus-31 (^{31}P) or hydrogen (^{1}H). However, the majority of tomographic images obtained in clinical MRI rely on the properties of hydrogen atoms, characterised by a single-proton nucleus. The hydrogen atom is preferred mainly due to two reasons: firstly, because it is naturally abundant *in-vivo* in the form of water (∼60% of the human body in weight) and lipid molecules and secondly, due to its large gyromagnetic moment [13]. For that reason, following sections will focus on the representation of a single ^{1}H proton with intrinsic characteristics of spin, mass and electric charge. The proton, commonly represented as a rotating

particle with mass and electric charge, presents a magnetic dipole moment (μ), which is related to the nuclear spin (I), as follows:

$$\mu = \gamma I \tag{2.1}$$

where γ represents the gyromagnetic ratio, a nucleus-specific property. Due to its magnetic properties, the nuclei behaviour can be influenced by placing them in an external magnetic field produced by the MRI scanner, conventionally notated as B_0. Additionally, μ undergoes precession (i.e. rotation around the main direction of B_0) at a Larmor angular frequency (ω_0):

$$\omega_0 = \gamma B_0 \tag{2.2}$$

where $\omega_0 = 2\pi f_0$. ^{1}H protons present a gyromagnetic ratio of $\gamma = 2.67 \times 10^8$ radians/s/T, that can also be described in the scalar form as $\frac{\gamma}{2\pi} = 42.58$ MHz/T. Thus, for example when imaging at 1.5 T, protons will present a constant Larmor frequency of about 63.87 MHz.

The spin of a ^{1}H proton is $\frac{1}{2}$, therefore the hydrogen proton can either assume a spin state parallel or anti-parallel to B_0, with distinct quantum energetic values, known as the Zeeman effect (Figure 2.1). In the presence of B_0, the number of spins distributed in the parallel or anti-parallel states is given by a Boltzmann distribution, described as follows:

$$\frac{N_{anti-parallel}}{N_{parallel}} = e^{\frac{-\gamma h B_0}{\kappa T}} \tag{2.3}$$

where κ stands for the Boltzmann constant, h for the Planck constant and T for temperature. When $T \neq 0K$ and $B_0 \neq 0$, a slightly higher amount of spins prefers the parallel orientation (characterised by a lower energetic state) and a slight net magnetisation builds up in the B_0 parallel direction.

The collective magnetic behaviour of the entire spins system can be described by a net magnetisation vector (M) with a magnitude given by:

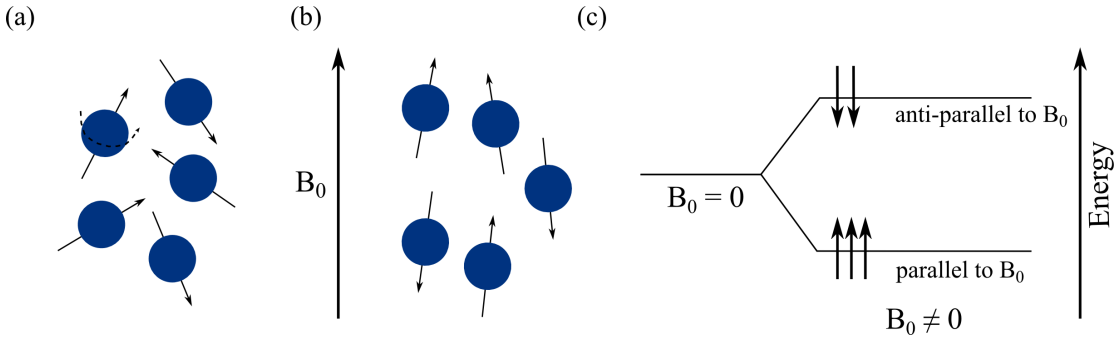


Figure 2.1: Schematic representation of ^{1}H protons with intrinsic angular momentum in (a) absent external magnetic field and (b) in the presence of an external magnetic field B_0 generated by the scanner magnet. (c) Zeeman distribution of spins in the parallel (low-energy state) and anti-parallel (high-energy state) direction dependent on an intrinsic energetic state.

$$M \approx \frac{h^2 \gamma^2 N_s}{4\kappa T} B_0 \tag{2.4}$$

where N_s represents the ^{1}H proton density. However, this magnitude of M is only in the order of a few μT and therefore, undetectable when in the same direction as the much stronger B_0.

2.2 Excitation and relaxation

The rate of change of the magnetisation vector M can be described by the well-known Bloch equation, given as:

$$\frac{dM}{dt} = \gamma M \times B(t) \tag{2.5}$$

where $M = \langle M_x, M_y, M_z \rangle$ and $B(t)$ encompasses all magnetic fields seen by M, thus including the main magnetic field (B_0), local field inhomogeneities in B_0 and an excitation field $B_1 = (B_1 cos\omega_0 t)_x - (B_1 sin\omega_0 t)_y$.

When in equilibrium, $M_z = M > 0$ and $M_{x,y} = 0$. However, when applying an excitation pulse (B_1) tuned to the specific Larmor frequency, spins are ener-

getically excited and M_z is tipped away from the direction of main field B_0. Due to the intrinsic precession undergone by M around the main direction of B_0 at the frequency of ω_0, this tipping of M_z can be described as a spiraling downwards motion of M precessing around the z-axis, as seen from a stationary (laboratory) frame of reference. The angle by which M_z is tilted to the transverse plane is known as the flip angle (α) and depends both on the amplitude and duration of the B_1 pulse. When B_1 is turned off, spins return to their equilibrium state by a process known as relaxation. The two relaxation mechanisms are known as longitudinal and transverse relaxation. Both relaxations effects are tissue-dependent and occur simultaneously.

Longitudinal or spin-lattice relaxation results from the dissipation of excess energy from the excited spins to their molecular surroundings (i.e. the lattice). This causes net magnetisation to regain magnitude in M_z. Longitudinal relaxation can be described exponentially by the following equation as a recovery of longitudinal magnetisation in the z-axis:

$$M_z(t) = M_{z,0}(1 - e^{\frac{-t}{T_1}}) \tag{2.6}$$

where T_1 is the time-constant corresponding to the time it takes for M_z to recover 63% of its original magnitude and $M_{z,0}$ is the initial magnetisation magnitude in equilibrium.

Immediately after the RF excitation pulse, spins are in phase-coherence. However, due to spin-spin interactions (also referred to as transverse relaxation), spins start to dephase relative to each other and net transverse magnetisation M_{xy} decays according to a time-constant known as T_2. An additional process contributes to a faster decay of M_{xy}, T'_2 relaxation, where $T'_2 \ll T_2$ and describes the dephasing of spins due to inhomogeneities in the magnetic field [13]. T_2^* relaxation as an exponential decay of M_{xy} due to these two phenomena can be described as follows:

$$M_{xy} = M_{xy,0} e^{\frac{-t}{T_2^*}} \tag{2.7}$$

where:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'} \tag{2.8}$$

Where T_2 is a time-constant corresponding to the decay of M_{xy} to 37% of its original magnitude ($M_{xy,0}$). Both the longitudinal and transverse relaxation curves are shown in Figure 2.2.

Relaxation effects can also be described by further expansion of the original Bloch equation and given as:

$$\begin{aligned} \frac{dM_x(t)}{dt} &= \gamma(M_y B_0 + M_z B_1 sin\omega_0 t) - \frac{M_x(t)}{T_2} \\ \frac{dM_y(t)}{dt} &= \gamma(M_z B_1 cos\omega_0 t - M_x B_0) - \frac{M_y(t)}{T_2} \\ \frac{dM_z(t)}{dt} &= \gamma(-M_x B_1 sin\omega_0 t - M_y B_1 cos\omega_0 t) - \frac{(M_z(t) - M_0)}{T_1} \end{aligned} \tag{2.9}$$

After excitation, the precessing transverse magnetization (M_{xy}) induces an alternating time-varying voltage in a receiver coil of the MRI scanner, known as the free induction decay signal (FID) in the laboratory reference, and the MR signal can be measured.

2.3 Magnetic Resonance Imaging

2.3.1 Signal formation

The previously described FID signal oscillates at the Larmor frequency and decays exponentially to zero due to the loss of phase coherence. In order to allow for spatial encoding, MR tomography most commonly makes use of an echo signal derived from manipulating the original FID. This echo is either generated by

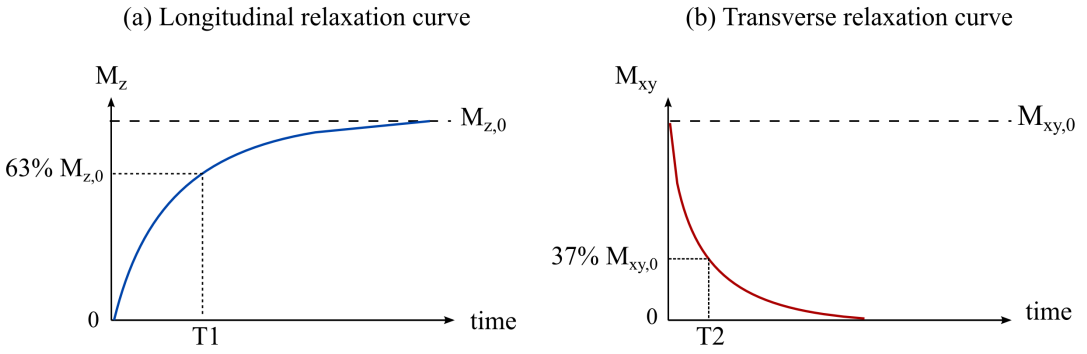


Figure 2.2: Schematic representation of spin-lattice (a) and spin-spin (b) relaxation curves that occur simultaneously over a certain period of time. After B_1 has been switched off, M_{xy} decays exponentially while magnetisation recovers in the z-axis and M_z recovers towards $M_{z,0}$.

the application of an extra RF pulse (spin echo SE) or the encoding gradient themselves (gradient echo). Gradient echo (GRE) is of particular interest for the work developed in this thesis and is detailed below.

GRE sequences rely on the use of gradients with opposing polarities to generate a gradient echo signal. Firstly, an excitation pulse of user-selected flip angle (usually less than 90°) is used to tip the magnetisation towards the xy-plane. After the excitation pulse, a dephasing gradient lobe is applied to force a rapid dephasing of the spins. Immediately following, a rephasing gradient (of opposing polarity) is applied to revert the effect and rephase the spins, leading to the formation of a gradient echo. Because the rephasing gradient only compensates the loss of phase due to the dephasing lobe, dephasing due to spin-spin relaxation and field inhomogeneities are unaffected. Therefore, the resulting gradient echo signal is proportional to the T_2^* decay [14]:

$$S_{GRE} \propto e^{\frac{-TE}{T_2^*}} \tag{2.10}$$

where TE is the echo time, corresponding to the time between the RF excitation pulse and the centre of the echo formation. In both SE and GRE sequences,

the time between two consecutive excitation RF pulses is known as repetition time (TR). Both TE and TR are user-defined parameters that allow controlling image contrast characteristics. GRE sequences have become particularly popular to perform fast MR imaging as they usually rely on smaller flip angles and much shorter TR values compared to SE sequences. Therefore, GRE total acquisition times can be greatly reduced.

2.3.2 k-space and signal encoding

In order to generate a MR tomographic image, it is not only necessary for the echo signal to be generated but equally important to be spatially encoded. That is, the signals must be distinguished when arising from different spatial locations within the imaged object. Spatial encoding is performed by means of spatially-variant magnetic field gradients (G), where field strength is varied linearly in accordance to spatial location, that is the field is stronger or weaker in different spatial locations. Because the precessional angular frequency (ω) is proportional to the total applied external magnetic fields (including additional spatially-variant gradients), it is possible to encode the spins' location in regards to their Larmor frequency:

$$\omega(x, y, z) = \gamma(B_0 + xG_x + yG_y + zG_z) \tag{2.11}$$

where (x, y, z) are the spatial coordinates of the spins. Data collected from these spins' spatially-varying frequencies is then organised in a matrix traditionally called k-space [15]. Each set of spatial frequencies corresponds to a position in the k-space matrix ($k_{x,y,z}$), which can be described as follows:

$$k_{x,y,z}(t) = \frac{\gamma}{2\pi} \int_0^t G_{x,y,z}(t)dt \tag{2.12}$$

When ignoring the relaxation effects, the MR signal *S(k)* equation can be described as the Fourier Transform of the detected magnetisation at the point r

$(m_0(r))$ [16]:

$$S(k) = \int m_0(r) e^{-i2\pi k(t).r} dr = \Im\{m_0\} \tag{2.13}$$

Where r represents the location coordinates and *k(t)* the k-space coordinates. This means that as t varies, *S(k)* traces a path along k-space according to *k(t)*. The image data can then be reconstructed by applying the inverse of the Fourier Transform, described as $\Im^{-1}\{S(k)\}$. Although this formula describes the relation between k-space and image data as a continuous signal, actual sampling of MR data is performed on a finite number of points (discrete), which will be further explored in Section 2.3.4.

As previously mentioned, for the complete formation of a tomographic image, the obtained MR signal also needs to be spatially encoded. Firstly, a cross-sectional region or slice must be chosen of certain thickness and location. Thus, a slice-encoding gradient (here referred as G_z) is applied at the time of RF excitation pulse. In practice, this means that each selected cross-section presents its unique combination of spatial frequencies ($\omega(z) = \gamma(B_0 + zG_z)$), depending on the spatial location of the spins along the z-direction (z). Therefore, by applying an RF excitation pulse centred at ω_0 and with a bandwidth of $\triangle\omega = \gamma \triangle zG_z$, it will only excite spins present in the corresponding slice.

Secondly, spatial resolving within the slice is accomplished by what is known as phase and frequency encoding. Phase encoding occurs by applying a gradient G_y. Application of the G_y gradient will once more affect the Larmor frequencies of the spins depending on their position along the gradient orientation. This means that protons located where the magnetic field gradient is stronger, will precess faster and gain a higher phase than protons located in the regions with a weaker magnetic field. After G_y is switched off, protons return to their original Larmor frequency but now with a distinct phase shift between them. This means that phase is now encoded spatially and varies along the direction of y. In a conventional 2D Cartesian grid, phase encoding allows for a shift in k_y and a N_y number of phase-

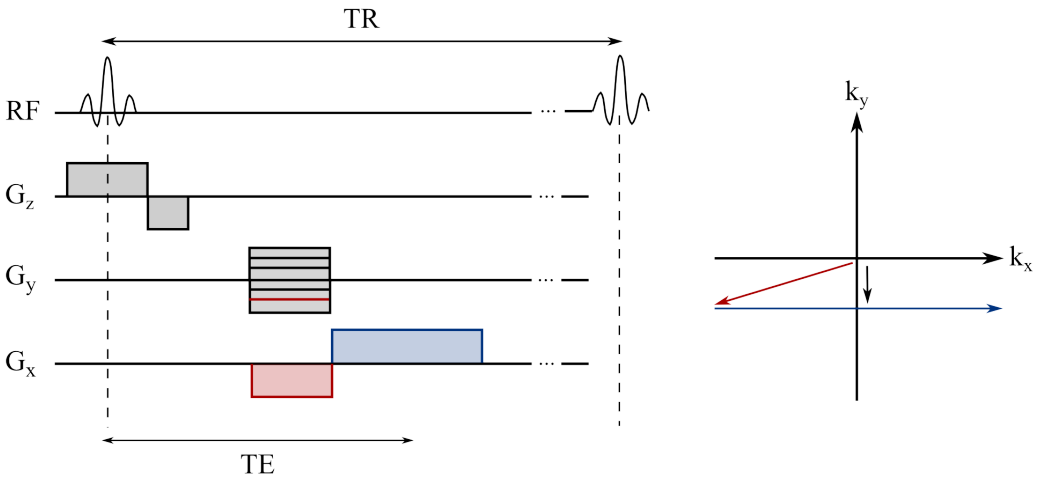


Figure 2.3: Example of sequence diagram for a basic 2D GRE sequence and corresponding temporal path on k-space. A RF pulse (B_1 field) and a slice-encoding gradient (G_z) are applied to limit spin excitation to a selected 2D slice. A phase-encoding gradient (G_y) is applied to encode information in the k_y direction. Lastly, a G_x gradient is applied along x allowing to encode information along the k_x direction (blue line in k-space). An echo formation occurs at the time of TE and the time between consecutive RF excitations is denominated as TR. Only one k-space line (blue line) is acquired in this TR, N_y repetitions are necessary of this scheme to acquire a full k-space.

encoding lines are necessary to sample an image. Therefore, G_y must be applied N_y times, each step having a different gradient amplitude, with a repetition time TR to obtain all 2D image information. Finally, a last step known as frequency encoding (notated as G_x) must be applied in order to fully encode (localise) the signal in the frequency encoding (readout) direction. This allows to encode the spins' precessional frequency to their spatial location. A GRE sequence example diagram can be seen in Figure 2.3.

2.3.3 GRE sequences: spoiled and SSFP sequences

As previously described, because GRE sequences mostly rely on short TR values and incomplete flipping of M_z to the transverse plane ($\alpha < 90$ degrees), each excitation pulse acts on the remaining available magnetisation from previous excitations. GRE sequences can from this point onwards be classified according to what happens to the remaining transverse magnetisation after each TR excitation.

One main group of GRE sequences is known as spoiled GRE sequences. These sequences purposely eliminate ("spoil") the remaining transverse magnetisation at the end of each TR excitation cycle. Spoiling of the transverse magnetisation at the end of each TR-cycle can be commonly performed with RF spoiling although other methods exist. Some of the names given to this type of sequences by various vendors might include Fast Low Angle Shot (FLASH), T1-Fast Field Echo (T1-FFE) or Spoiled Gradient Echo (SPGR). A T1-FFE sequence will be frequently used in following experimental chapters. Magnetisation given by a spoiled-GRE sequence can be described as:

$$M_{spoiled} = M_0 \frac{sin(\alpha)(1 - e^{\frac{-TR}{T_1}})}{1 - cos(\alpha)e^{\frac{-TR}{T_1}}} e^{\frac{-TE}{T_2^*}} \tag{2.14}$$

The flip angle that maximizes $M_{spoiled}$ is given by $\alpha_e = cos^{-1}(e^{\frac{-TR}{T_1}})$ and is known as the Ernst angle.

On the other hand, after a certain number of TR repetitions and specific conditions; such as a constant TR, a total gradient area on any axis within each TR interval equal to zero and when no magnetisation spoiling is applied, a steady state magnetisation is achieved. This family of GRE sequences are most commonly called Steady State Free Precession (SSFP) sequences [17]. A particular case of SSFP sequences are the balanced SSFP (bSSFP) sequences, also referred to as FIESTA, balanced FFE or TrueFISP [18]. In this case, in each TR, the applied gradient is compensated by a gradient of opposing polarity in order to rephase magnetisation in all three axes.

The signal from a bSSFP sequence can be described as [16, 18]:

$$M_{bSSFP} = M_0 \frac{(1 - E_1)sin(\alpha)}{1 - (E_1 - E_2)cos(\alpha) - E_1 E_2} e^{\frac{-TE}{T_2}} \tag{2.15}$$

Where $E_1 = e^{\frac{-TR}{T_1}}$ and $E_2 = e^{\frac{-TR}{T_2}}$. If considering the assumption $TR << T_1, T_2$, thus the previous equation may be simplified as $E_{1,2} \approx 1 - \frac{TR}{T_{1,2}}$, becoming:

$$M_{bSSFP} = M_0 \frac{sin(\alpha)}{(T_1/T_2)(1 - cos(\alpha)) + (1 + cos(\alpha))} e^{\frac{-TE}{T_2}} \tag{2.16}$$

The optimal flip angle is given by:

$$cos(\alpha) = \frac{\frac{T_1}{T_2} - 1}{\frac{T_1}{T_2} + 1} \tag{2.17}$$

Therefore, the bSSFP signal is given by:

$$M_{bSSFP} = \frac{1}{2} M_0 \sqrt{\frac{T_2}{T_1}} \tag{2.18}$$

Equation 2.18 not only demonstrates that bSSFP images present an image contrast based on the $\frac{T_2}{T_1}$ ratio, but also present a high signal performance (up to 50% of M_0) thus making high signal-to-noise (SNR) an intrinsic characteristic of bSSFP sequences. Differences in sequence diagram between an SSFP and bSSFP are shown in Figure 2.4.

2.3.4 Sampling of k-space

As previously mentioned in Section 2.3.2, the continuous signal *S(k)* detected in MRI is not measured continuously but in fact as a finite number of discrete k-space data points at a distance given by Δ_{kr}. The discrete sampling of $S(k_n)$ is then expressed as the multiplication with a Dirac comb function ($D_c(k)$):

$$S(k_n) = S(k)D_c(k) = S(k) \sum_{-n_{max}}^{+n_{max}} \delta(k - n\Delta_{kr}) \tag{2.19}$$

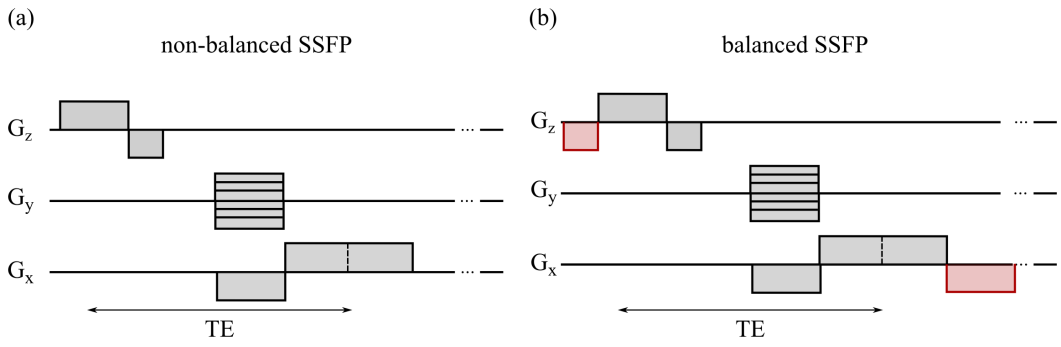


Figure 2.4: Sequence diagrams highlighting the main differences between (a) a non-balanced SSFP and (b) a balanced SSFP sequences. While in the case of non-balanced sequences, the Gx gradient is not balanced (different gradient areas in the positive and negative polarities), magnetisation will dephase after readout at the end of each TR cycle. In the case of balanced SSFP, an additional lobe of the gradient is applied in order to balance gradient areas in both polarities, leading to a single magnetisation vector at the end of TR. Image adapted from [18].

This means that a final reconstructed image can be obtained from a limited number of data samples, although with replicas of the signal shifted at intervals of $\frac{1}{\Delta_{kr}}$ in image space. The requirement to define the necessary sampling distance to obtain an aliasing-free image is given by the Nyquist sampling theorem. Applying this principle to MR imaging and in order to obtain an image with resolution $\Delta_{x,y}$ and field-of-view ($FOV_{x,y}$), k-space must be sampled at discrete intervals of Δ_{kr} to avoid aliasing (superimposition of signal replicas), given as:

$$\Delta_{kr} = \frac{1}{FOV_{x,y}} \tag{2.20}$$

Also, for an image with a chosen spatial resolution, the maximum distance reached in k-space ($k_{x,y}^{max}$) and thus the maximum number of sampled k-space points (n_{max}) are defined as:

$$k_{x,y}^{max} = \frac{1}{2\Delta_{x,y}}$$
$$n_{max} = \frac{k_{x,y}^{max}}{\Delta_{kr}} \tag{2.21}$$

Most conventionally, k-space sampling is performed in a line-by-line scheme known as Cartesian sampling. However, k-space sampling is not limited to a Cartesian coordinates scheme and can in fact be sampled using any arbitrary trajectory with corresponding gradient waveforms. Two of the most commonly used trajectories, and of particular interest to the experimental work in this thesis, are the radial and spiral trajectories. Both sampling schemes are explored in the following subsections.

2.3.4.1 Non-Cartesian sampling: Radial

As shown in Figure 2.5, radial sampling consists of sampling k-space along radial lines centred at $k_{x,y} = 0, 0$. Since each line of radially sampled k-space is equivalent to sampling the centre line (no phase encoding gradient) of a Cartesian k-space, all other radial phase-encoding lines (also known as spokes or radial projections) can be derived from this line using an angular rotation given by ϕ. Therefore, gradients G_x and G_y can be described as:

$$G_x = G_0 cos(\phi) \tag{2.22}$$

$$G_y = G_0 sin(\phi) \tag{2.23}$$

Where G_0 denotes the gradient waveform used to sample the central k-space phase-encoding line. Conventional 2D radial trajectory implementation follows a uniform angular sampling pattern in either $0 < \phi < 2\pi$ or $0 < \phi < \pi$. Thus, for a set number of radial spokes (N_{shots}), the rotational angle ϕ in π is defined as:

$$\Delta\phi = \frac{\pi}{N_{shots}} \tag{2.24}$$

In addition, angular rotation can also be performed with other slight variations, such as the Golden Angle ratio (GA) method. Instead of a conventional uniform sampling, ϕ_{GA} is defined in degrees by:

$$\phi_{GA} = 180/\frac{\sqrt{5}+1}{2} \approx 111.25 \tag{2.25}$$

In order to fulfill the Nyquist criterion and avoid aliasing artefacts, the distance between consecutive data points in the angular direction (between consecutive radial spokes) must be the same as the distance between data points in the readout direction Δk_r. Thus, the number of radial projections (N_{shots}) is given by taking into consideration the number of phase-encoding lines (i.e. acquisition matrix) in an equivalent Cartesian k-space given by N_r, such as:

$$N_{shots} = \frac{\pi}{2} N_r \tag{2.26}$$

Because of the uniform sampling characteristic of these non-Cartesian trajectories and in order to fulfill the Nyquist criterion at the edges of k-space, the centre portion is always oversampled. This means that compared to a Cartesian case of equal FOV and spatial resolution, the radial sampling requires $\sim 57\%$ more samples for a fully-sampled k-space.

2.3.4.2 Non-Cartesian sampling: Spiral

Spiral [19] trajectories consist of a spiralling sampling pattern of k-space covering data points both in the centre and outer k-space. The spiral k-space trajectory relies on the use of sinusoidal readout gradients waveforms, that are both amplitude and frequency modulated. As most spirals begin k-space sampling at the centre $k_{x,y} = 0, 0$; amplitude of the waveform is zero at the beginning of readout, gradually building in amplitude at a constant slew rate (S_r) until a constant amplitude (G_0) is reached. After G_0 is reached, amplitude remains constant while waveform frequency decreases. Maximum slew rate and amplitude values are intrinsically

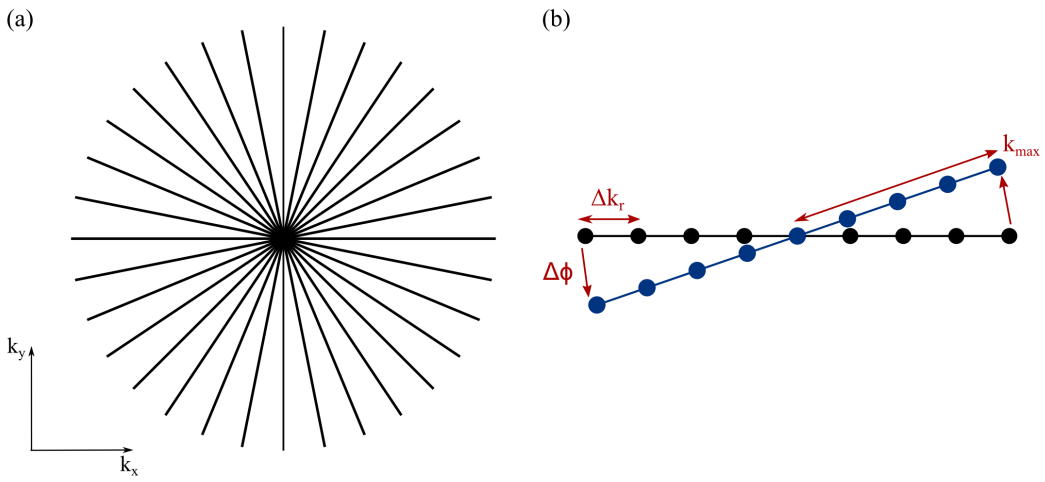


Figure 2.5: 2D Uniform radial trajectory described in (a) k-space (k_x,k_y) coordinates. (b) All radial projections (blue line) can be described as an angular rotation from the Cartesian centre k-space phase-encoding line (black line). Distance between sampling points is described as Δk_r and k_{max} gives the maximum k-space distance achieved.

dependent on the gradient hardware characteristics.

Spiral trajectories can be mathematically described with the use of polar coordinates (k, Θ) that relate to the Cartesian grid (k_x, k_y) as follows:

$$k_x = kcos(\Theta) \tag{2.27}$$

$$k_y = ksin(\Theta) \tag{2.28}$$

A simple spiral design in which the azimuthal angle Θ is proportional (by a constant λ) to the radius k, is known as an Archimedean spiral and described as:

$$k(t) = \lambda\Theta(t) \tag{2.29}$$

Spiral acquisitions can sample k-space either in one single shot (one single TR excitation) or in multiple excitations (i.e. spiral interleaves in multiple TR repetitions). Similarly to radial trajectories, spiral acquisitions usually use the

same trajectory for all spiral interleaves (N_{shots}), where each interleaf is rotated by an angle $\phi = \frac{2\pi}{N_{shots}}$. The radial distance between consecutive spiral interleaves is given by $\Delta_k = \frac{2\pi\lambda}{N_{shots}}$. Aliasing artefacts in spiral trajectories due to radial under-sampling (i.e. distance between consecutive interleaves) or azimuthal under-sampling (i.e. distance between data points in each interleave) are independent. To obtain an aliasing-free image in regards to the radial direction and following equation 2.20, then it must be given that [16]:

$$\lambda = \frac{N_{shots}}{2\pi FOV} \tag{2.30}$$

In addition, desired gradient amplitude is defined by the Nyquist criterion at the azimuthal direction [16]:

$$G_0 = \frac{4\pi BW}{FOV\gamma} \tag{2.31}$$

Where BW is given in units of s^{-1} and the gyromagnetic ratio (γ) is given in radians/s/T. One of the main advantages of spiral sampling is the time-efficient coverage of k-space, sampling centre and outer k-space data points within one excitation. For this reason, spirals have been frequently used in fast imaging such as real-time MRI, a focus of this thesis.

2.4 Image Reconstruction

Image reconstruction consists of the process of obtaining a final image of the selected object from the measured MR signal. As previously described, data sampled along a conventional Cartesian grid ($N \times N$) can be directly reconstructed to image space with a discrete inverse fast Fourier Transform (iFFT). This allows a fast calculation of the reconstructed image and reduces the number of operations from $\propto N^2$ to N×log(N). However, non-Cartesian trajectories do not sample data points in an equidistant grid and thus a direct application of iFFT is no longer possible.

2.4.1 Non-Cartesian reconstruction: gridding

A common approach to reconstruct non-Cartesian data is known as gridding [20]. The idea behind gridding is to interpolate the non-uniformly sampled data points onto a Cartesian uniform grid so that a FFT can be directly applied afterwards. There are many advantages to this technique, from being fairly simple to implement to being computationally more efficient than iterative reconstruction schemes. In addition, gridding reconstruction methods can be applied to any form of non-Cartesian k-space, independently of the specific trajectory (e.g. radial, spiral, propeller, etc.).

A gridding reconstruction pipeline usually includes the following steps: (i) application of a density compensation function (DCF), (ii) convolution of the k-space data with a chosen gridding kernel (iii) oversampling to Cartesian grid (iv) application of iFFT, (v) application of de-apodisation function and (vi) crop final image.

In non-Cartesian trajectories, data points in the centre of k-space are usually sampled more frequently than those at the periphery, resulting in an over-sampling of low spatial frequencies. To correct for this, a DCF must be applied to weight data points differently and allow for an equal final representation of spatial frequencies. The DCF function can be applied to k-space points before the interpolation step or to the later Cartesian re-gridded data points. Several density compensation methods have been described [21]. In the case of angular uniformity as in radial trajectories, an analytic density compensation function can be derived from using a conventional Ram-Lak (or ramp) filter:

$$\begin{aligned} D_{ramp}(k) &= \frac{|k|}{N_{spokes}} if |k| \neq 0 \\ &= \frac{1}{2N_{spokes}} if |k| = 0 \end{aligned} \tag{2.32}$$

On the other hand, more complex trajectories such as spirals, can be compensated with other methods such as the Voronoi diagram, where the area around

each k-space sampling point is calculated. Effect of the application of the DCF in simulated radial k-space can be seen in Figure 2.6.

Secondly, each density compensated k-space point is convolved with an interpolation kernel and resampled onto a Cartesian grid. After an extensive comparison of kernel windows for optimal interpolation, Jackson et. al suggested a Kaiser-Bessel window (W_{KB}) as providing optimal image quality [22]. The Kaiser-Bessel window is described as follows:

$$\begin{aligned} W_{KB}(d) &= \frac{1}{L} I_0(\beta\sqrt{1-(\frac{2d}{L})^2}) if |d| < \frac{L}{2} \\ &= 0, otherwise \end{aligned} \tag{2.33}$$

where L represents the Kernel width, I_0 is the zero-order Bessel function and β describes the kernel shape [23]. Optimal selection of the shape of the window has also been described by Beatty et al. [23]. After convolution with the Kaiser-Bessel window, data points are then resampled onto a Cartesian grid. However, due to the side lobes of the gridding kernel, aliasing may arise within the imaged field-of-view. In order to avoid this artefact, oversampling of the Cartesian grid is usually performed by an oversampling factor, such as 1.5 or 2. This allows to shift the aliased side lobes away from the object so that they are outside the cropped field-of-view of the final produced image.

Following data interpolation and sampling onto a uniform Cartesian grid, an iFFT can be applied to reconstruct the present Cartesian k-space. Lastly, due to the convolution with a finite kernel window, a corresponding modulation with the Fourier transform of the kernel is induced. Thus, compensation of this effect (de-apodisation) can be achieved by dividing the resulting image by the Fourier transform of the Kaiser-Bessel window:

$$\Im\{W_{KB}\} = \frac{sin(\sqrt{(\pi Lx)^2 - \beta^2})}{\sqrt{(\pi Lx)^2 - \beta^2}} \tag{2.34}$$

Similarly to the gridding method, a non-uniform FFT (NUFFT) has also been

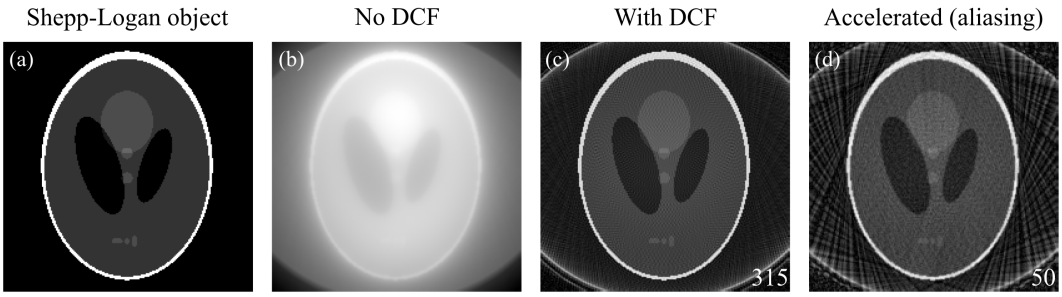


Figure 2.6: Reconstruction of radial k-space and effect of density compensation function. (a) Original Shepp-Logan object used for simulated data. (b) Image reconstructed from simulated radial data without the application of a density compensation function (DCF). (c) Image reconstructed from radial data using 315 radial projections (200 readout points each) with application of DCF, removing low-frequency enhancement visible in (b). (d) Image reconstructed from accelerated radial data (only 50 radial projections), where typical radial streaking artefacts across the image are visible due to aliasing.

described and optimised for reconstruction of non-uniform data. This has been made widely available by Fessler (Image Reconstruction open-source toolbox) [24] and was recurrently used in the subsequent experimental chapters.

2.5 Accelerating data acquisition

The investigation of methods to accelerate MR data acquisition has always been of interest in the MR field due to the longer acquisition times when compared to other imaging modalities. Considering a standard GRE sequence, the total acquisition time of a single 2D slice will directly depend on two main parameters: the time it takes to acquire one phase encoding step (TR) and the total number of phase-encoding steps. Therefore, different approaches can be considered when trying to reduce total scan time; to either acquire the same amount of k-space data faster (shorter TR), acquire multiple phase-encoding lines per TR (e.g. echo-train sequences) or acquire less k-space data (reduce phase-encoding lines). The use

of shorter TR values is commonly used in GRE sequences, furthermore, shorter TR values are intrinsically limited by hardware and physiological constraints (e.g. increased heating and likelihood of peripheral nerve stimulation [25]).

On the other hand, total scan time can be reduced by skipping phase-encoding lines in k-space, that is to increase Δ_{kr}. However, as explained in Section 2.3.4, the distance between k-space phase-encoding lines (Δ_{kr}) directly influences FOV and presence of aliasing artefacts (Figure 2.7). While aliasing in Cartesian k-space presents a coherent behaviour and will generate the replication or folding-over of the imaged object onto itself in the phase-encoding direction, the same is not true for non-Cartesian sampling. With non-Cartesian k-space, aliasing artefacts present a less coherent nature and for example, if using radial trajectories, will appear as streaking artefacts across the image. Opposite to Cartesian aliasing, this means that overall characteristics of the object are still identifiable and moderate degrees of under-sampling in non-Cartesian imaging can usually still be tolerated. However, if high degrees of acceleration are intended (i.e. many phase-encoding lines are skipped), aliasing artefacts may severely hamper image quality thus making the need for methodologies that decrease the effect of aliasing in final image quality.

2.5.1 Parallel imaging

One of the most commonly known methods to accelerate data acquisition in MRI is the principle of Parallel Imaging, extensively covered in review publications such as [26–28]. The basic principle behind parallel imaging is that aliasing (due to under-sampling of k-space) can be unfolded through the use of additional encoding information provided by a multi-channel receiver coil. Each coil element within the array presents a unique coil sensitivity map, i.e. each coil is more sensitive to the signal arising in neighbouring tissues. Thus, each coil can provide additional encoding information about the acquired signal.

Many parallel imaging techniques have been described, usually divided within

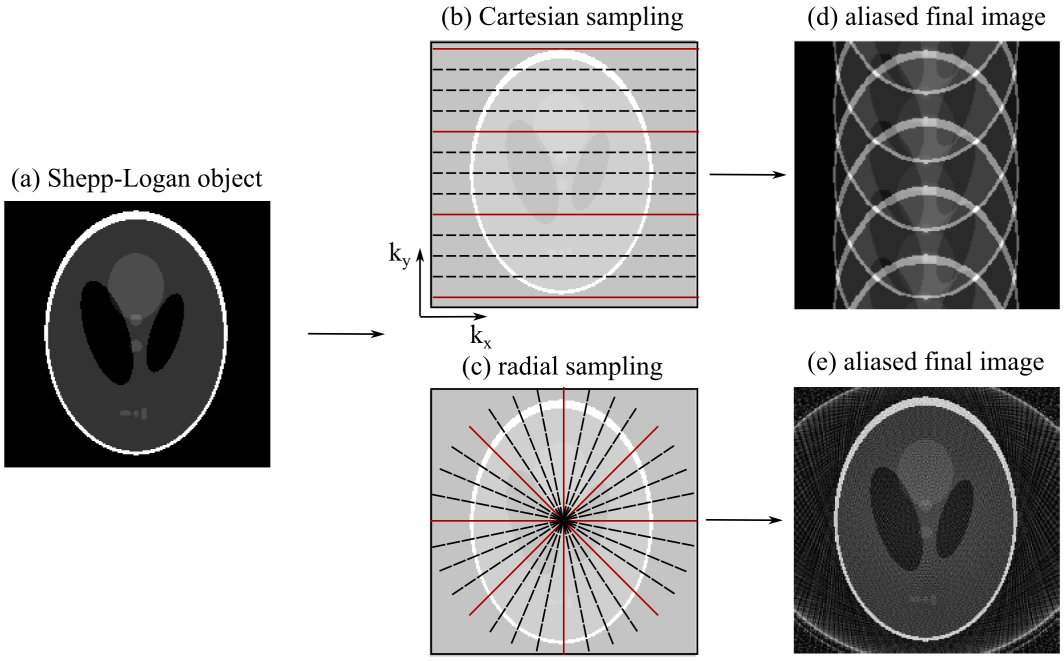


Figure 2.7: Effect of under-sampling Cartesian and non-Cartesian k-space by a factor 4. (a) Original Shepp-Logan object used for simulated data. (b) Example of under-sampling (reducing acquired phase-encoding lines, only acquiring red lines) in Cartesian k-space. (c) Example of under-sampling (reducing acquired radial projections, only acquiring red lines) in radial k-space. (d) Effect of under-sampled Cartesian k-space in final reconstructed image results in aliasing and replicas of the object along the phase-encoding direction. (e) Effect of under-sampled radial k-space results in less coherent aliasing artefacts, usually translated by radial streaking across the image although with clear definition of the original object shape.

two main families, by either trying to resolve aliasing in image space (SENSitivity Encoding SENSE [29]) or by estimating missing data in k-space (Generalized Autocalibrating Partially Parallel Acquisitions GRAPPA [30]).

As a commercial form of SENSE is used in the experimental Chapter 4, a brief account of the principles of SENSE is presented. In SENSE, data is under-sampled by a factor R, thus increasing the distance between adjacent phase-encoding lines. This results in an aliased final image where signals arising from R pixels are wrapped into one location. SENSE then uses the additional encoding information from each coil to unfold this aliased image. Thus, for an example with an acceleration factor of R = 2, at a pixel location (x, y) in the aliased image, the signal resulting from two example coil elements (a_1 and a_2) is given by [27]:

$$a_1(x,y) = C_1(x,y)u(x,y) + C_1(x, y + \frac{FOV}{2})u(x, y + \frac{FOV}{2}) \tag{2.35}$$

$$a_2(x,y) = C_2(x,y)u(x,y) + C_2(x, y + \frac{FOV}{2})u(x, y + \frac{FOV}{2}) \tag{2.36}$$

where C_1 and C_2 denote the coil sensitivities for the two coil elements and u the signal from the object at the two locations (x, y) and $\langle(x, y) + \frac{FOV}{2}\rangle$. By considering the two equations simultaneously and assuming the knowledge of coil sensitivities (C_1,C_2,etc.), it is possible to solve the equation system and unfold the aliased image. This problem can be generalised for an arbitrary number of coil elements and acceleration factor R using the following matrix form:

$$a = Cu \tag{2.37}$$

where C is a matrix of $n_c \times R$, corresponding to the number of coil elements and the number of folded pixels. In order to be able to solve this system of equations and unfold the image, it is necessary to acquire at least as many coil elements as

the used acceleration factor R. The unknown image intensity u can then be solved using the Moore-Penrose inverse of the matrix C:

$$u = (C^H C)^{-1} C^H a \tag{2.38}$$

While SENSE is said to be an image-based method, where the unfolding of the aliased pixel intensities is performed in image space, GRAPPA is a k-space based method. Because the principles of GRAPPA and its application to non-Cartesian imaging are the basis of some experimental work in this thesis, a full description of the technique is provided in Chapter 6. A well-known disadvantage of parallel imaging, whether image or k-space based, is the reduction of SNR, dependent on the acceleration factor R and a so-called g-factor [27]:

$$SNR_{accel} = \frac{SNR_{full}}{g\sqrt{R}} \tag{2.39}$$

where SNR_{accel} is the SNR of the accelerated data, SNR_{full} is of the corresponding SNR of the fully sampled data and g depends on several factors such as coil geometry.

2.5.2 View-sharing

Another technique referred to during this experimental work is the application of a sliding window (or view-sharing techniques). View-sharing can be simply described as a temporal filter that can slide through the acquired data in the temporal dimension and reconstruct a new set of dynamic frames. Considering a case of radial sampling where 100 radial projections are acquired at a frame rate of N frames per second (fps), a sliding window of 2 can be considered by reconstructing every 50 consecutive radial projections in order to obtain a final frame rate of 2N fps. View-sharing has been frequently described in the acceleration of temporal data in real-time MRI of speech [7, 8, 31]. View-sharing is used in experimental Chapter 5 and a more comprehensive explanation is given in this chapter.

Chapter 3

Speech and Imaging the Upper Vocal Tract

The complexity of human speech has always been of great interest to many in the fields of clinical assessment of speech, linguistics, cognitive science and others. With the increased use of imaging techniques such as x-ray videofluoroscopy and more recently MRI, visualisation in real-time of this complex and dynamic co-articulation of the vocal articulators became possible. In addition, questions such as "What does upper vocal tract imaging bring to speech assessment?" and "Can MRI bring new insights to speech research and in what way?", recently became of interest. This chapter presents the background in speech production and assessment necessary to understand the experimental work undertaken in this thesis. A description on the mechanisms of speech, speech disorders and MRI of the upper vocal tract is given.

3.1 Mechanisms of Human Speech

Human speech is a complex process that relies on the accurate and timely co-articulation of numerous organs such as: the lips, tongue, palate, pharynx and vocal cords, among others (Figure 3.1).

The airflow created by the lower respiratory organs (lungs) travels through the bronchial tubes into the upper portion of the respiratory tract. There, the ascending airflow passes the larynx and is forced through the narrow passage between the vocal cords (or folds), making them vibrate and produce sound. By manipulating length, strain and separation of the vocal cords, all major aspects of speech production (such as volume, frequency and pitch) can be controlled [3].

Velopharyngeal closure, a necessary mechanism during speech, consists in the controlled closure of the velopharyngeal port. This occurs with the posterior-superior elevation of the velum (also known as soft palate) and approximation of the pharyngeal walls. Additionally, other aspects such as shaping of the tongue and pursing of the lips also influence the final sound produced.

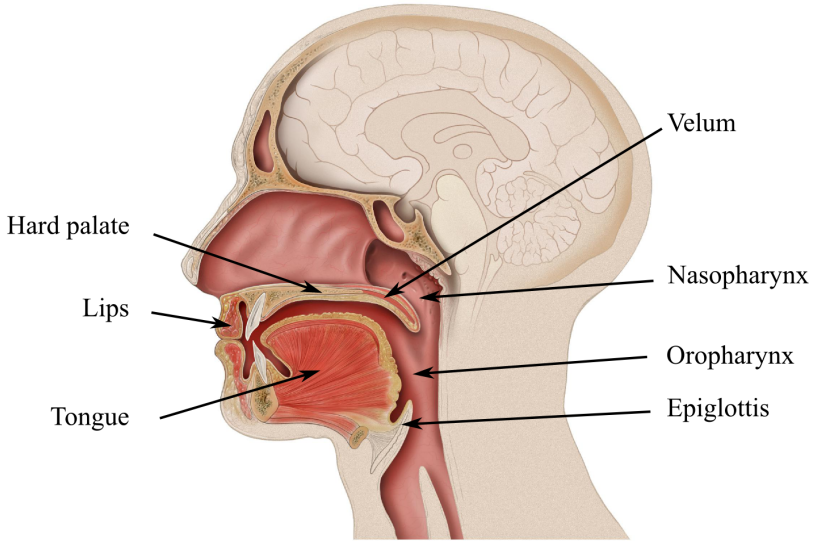


Figure 3.1: Schematic representation of the upper vocal tract structures relevant to speech production, such as the hard palate, lips, tongue, velum (or soft palate), pharynx and epiglottis. Image adapted from [32].

3.2 Disorders of speech: orofacial clefts and velopharyngeal insufficiency

A number of disorders, both due to acquired or inherited causes, may affect the performance of the vocal tract articulators and hamper the production of natural speech. These may include but are not limited to: vocal tract cancer, dysphagia, neurological disorders and orofacial clefts. In this thesis, orofacial clefts, and thus the imaging of velopharyngeal motion, are the main clinical focus.

Orofacial clefts are a set of congenital deformities that may encompass disorders of the lips, velum, pharynx and/or hard palate. Orofacial clefts are characterised by the partial or complete fissure of the upper lip, velum and/or hard palate. Affected newborns may present isolated cleft lip, isolated cleft palate or a range of variations in-between (Figure 3.2). An international study in 2011 from a total registry of 7.5 million births, reported a prevalence of orofacial clefts of about 9.92 per 10,000 births worldwide [1]. However, prevalence of orofacial clefts greatly varies depending on factors such as ethnicity, genetic predisposition and environmental factors (e.g. smoking during pregnancy) [33]. Besides difficulties in breathing, feeding and increased risk of ear infection, cleft patients commonly present disorders of speech.

Velopharyngeal insufficiency (VPI), common in orofacial clefts, is a speech impairment resulting from the incomplete closure of the velopharyngeal port. As a result, the air escapes through the nasal cavity during phonation and subjects present impaired speech. VPI patients commonly present hypernasality, impaired consonant production, abnormal nasal airflow, among other symptoms [34].

3.3 Clinical assessment of speech

Clinical assessment of VPI relies on a multi-disciplinary approach, including both the speech therapists perceptual evaluation, imaging techniques and others [35].

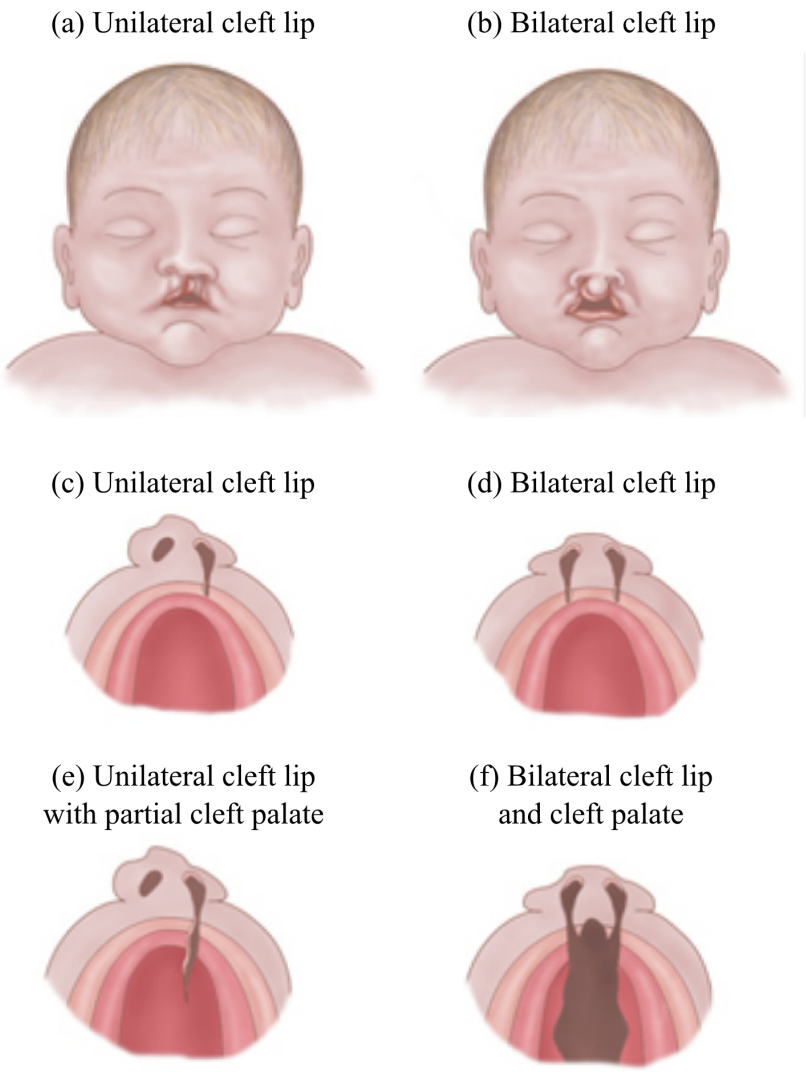


Figure 3.2: Example of orofacial clefts characterised by a partial or complete fissure of the lip, hard palate and/or velum that may be present both unilaterally or bilaterally. [34]

Clinical assessment of VPI most frequently starts with a perceptual evaluation by a speech and language therapist. There, the therapist may assess resonance and articulation based on a standardised speech task and/or repetition of sounds [2].

Additionally, imaging techniques have gained greater relevance in supplementing perceptual assessments. Direct assessment techniques allow for a direct visualisation of the vocal articulators during speech and can provide additional guidance in treatment planning. Two of the most commonly used imaging techniques in VPI assessment are x-ray videofluoroscopy and/or nasendoscopy [34].

X-ray videofluoroscopy is a non-invasive imaging technique where fluoroscopic images are consecutively acquired while the subject produces a selected speech task. Fluoroscopic images can be acquired in multiple planes (e.g. lateral, frontal and/or face planes). This allows for the measurement of anatomical features, such as velum thickness, velum length and pharynx depth, and assessment of velar motion, extent and timing of closure. Images are usually acquired at a rate of around 15 frames per second (fps). Although x-ray videofluoroscopy provides adequate visualisation of the hard palate and pharyngeal walls; soft tissue contrast, such as the velum, is relatively poor. To improve contrast, a suspension of barium is commonly applied to the vocal tract area by nasal aspiration [9]. However, this renders the procedure more unpleasant and a major constraint in younger patients [9]. Furthermore, one major disadvantage of videofluoroscopy concerns the repetitive exposure to ionising radiation. Swallowing videofluoroscopy studies reported a mean effective dose between 0.14 mSv for paediatrics between 1 and 3 years old [36] and up to 0.85 mSv for adults [37]. Since the absorbed dose is proportional to exposure time, selected speech tasks must be kept between 80 seconds and 2 minutes, and acquisition of multiple view planes kept to a necessary minimum.

Nasendoscopy consists of the passage of a fibre-optic scope transnasally into the nasopharynx, providing an *en-face* view of the velopharyngeal port. It allows for a good visualisation of velum motion, the presence of Passavants ridge and

configuration of the adenoids. The adenoids, also known as nasopharyngeal tonsils, are a mass of lymphatic tissue present in the roof of the nasopharynx, where the nasal cavity joins the pharyngeal space. Enlarged adenoids (common in infancy) may cause hampered nasal breathing as well as speech production [34]. However, when using nasendoscopy, a quantitative analysis of the velopharyngeal structures is limited and evaluation is mainly qualitative [2]. One advantage of nasendoscopy is that it does not involve the use of ionising radiation and thus, longer or repetition of speech tasks are possible. On the other hand, introduction of the nasal scope is rather invasive and requires full patient cooperation. Younger patients or those with a deviated nasal septum might require local anaesthetic [34].

Currently, both videofluoroscopy and nasendoscopy are commonly performed at several stages of assessment of a VPI patient. Imaging can be helpful both to assess the need for surgery, as well as planning, and to determine the course of further speech therapy. However, there is no true "gold standard" procedure and choice of imaging protocol highly depends on the patients age, type and severity of insufficiency, as well as clinical centre. For example in the UK, a recent report by Sell et al. has suggested that clinical protocols to assess VPI vary depending on cleft assessment clinic; as 100% of the 16 enquired use lateral videofluoroscopy and about 70% also use nasendoscopy [34]. Some efforts to standardise the clinical practice were made in 1990 by an international workgroup [38], by introducing a standard rating scale for the reporting and assessment of velopharyngeal disorders.

Limitations of videofluoroscopy and nasendoscopy have strongly fomented the interest in using MRI for the investigation of speech and imaging the upper vocal tract, as summarised in a review of the field [3]. MRI provides tomographic images with improved soft tissue contrast, ideal for vocal tract imaging, in multiple image planes.

3.4 Real-time MRI of speech: an overview

Limitations on previously described techniques used in standard assessment of speech impairments have strongly supported the use of real-time MRI in speech research and clinical translation. Many studies have used MRI to perform static imaging of the vocal tract while the subject is either silent or in sustained phonation [3]. These provide high-resolution anatomical information of the vocal tract articulators, and underlying muscles, until then limited with previous techniques. However, speech does not present a static behaviour but relies on the continuous interaction between vocal articulators. Thus, static MRI of the upper vocal tract, although valuable, can only provide limited information on the production of speech, and dynamic MRI methodologies are essential.

3.4.1 Upper vocal tract imaging requirements and considerations

Dynamic imaging of the upper vocal tract during speech presents distinct imaging requirements, primarily dependent on speech task, cohort population and/or speech condition at hand [39].

One of the main challenges when imaging velopharyngeal motion is the rapid transition of the velum between relaxed and elevated positions during closure. One of the first studies to focus on this issue, performing velar timing measurements in healthy subjects with x-ray videofluoroscopy, reported velum motion from relaxed position to consecutive relaxed position to last between 50 to 150 ms [40]. When testing different speaking rates (slow, normal and fast), Kuehn [40] observed that a healthy velopharyngeal mechanism adequately compensates for quicker speaking rates by shortening both the velar timing and velar displacement. Studies on assessing velopharyngeal closure with MRI have described a wide range of temporal resolutions, with images acquired every 50 ms to 200 ms [4, 6, 9]. However, and at the risk of missing velum closure events due to insufficient frame rate, Lingala

et al. recommends a target zone of around 100 ms [39]. Additionally, preliminary experiments where velum timing and closure were assessed with rt-MRI protocols, have recently suggested a median temporal resolution of about 7.5 fps (133 ms) to 10 fps (100 ms) for suggested speech tasks [41]. In summary, temporal resolution plays a key part when assessing velopharyngeal closure and careful consideration should be given in this work when selecting acquisition frame rate.

However, MRI acquisition is a relatively slow imaging technique (for example compared to videofluoroscopy studies) and adequate temporal sampling of vocal tract articulators can be challenging. As previously explained in Chapter 2, acceleration of data acquisition is limited by many constraints that are intrinsic to MRI. Thus, temporal resolution is limited by an inevitable trade-off in SNR, spatial resolution and/or overall image quality.

To overcome this issue, several studies have suggested the use of gated techniques, where a segment of k-space is acquired during a single repetition of the selected speech task. This then requires the subject to repeat the selected speech task as many times as necessary until all segments of k-space have been acquired. Although high frame rates have been reported [3], gated techniques present some limitations, as they require the accurate and consistent reproduction of the speech sample multiple times per acquisition. Unlike gated techniques, real-time MRI consists on the continuous acquisition of image data and may be more suitable to image natural speech production. Additionally, real-time imaging coupled with on-the-fly reconstruction could also in the future allow for an interactive speech exam led by a speech and language therapist. For this reason, real-time MRI of speech requires the investigation of particularly fast image acquisition. Imaging the upper vocal tract during speech with real-time MRI is the main focus of this experimental work, thus all work mentioned in the following sections of this chapter solely refers to real-time MRI techniques.

Contrary to temporal frame rate, not much work has been reported on a minimum in-plane spatial resolution to be considered. However, careful consideration

on the chosen voxel size should still be taken, in particular in paediatric cohorts where the velum is usually of smaller dimensions. Previous studies [4,6,9,42] have reported values between 3.6 (1.9×1.9) mm^2 and 5.3 (3.3×1.6) mm^2, however in a variety of settings (e.g. adult *vs.* children populations or type of speech assessment). In addition, groups interested in the study of velopharyngeal gap should be particularly wary of voxel size [39]. Single-slice mid-sagittal protocols have been most frequently used to image velum motion and velopharyngeal closure, with slice thickness ranging between 6 and 10 mm [4, 6, 9, 42].

Furthermore, dynamic imaging of the upper vocal tract is exceptionally challenging due to the surrounding air cavities around the vocal tract structures. Differences in magnetic susceptibility between air and tissue of the upper airway create local field inhomogeneities at these air-tissue boundaries and may result in detrimental image artefacts (see for example Figure 3.3). Image distortion due to off-resonance is particularly pronounced when imaging at higher field strengths and/or with longer readout time sequences (e.g. EPI or spirals). Alternative methodologies can be used to compensate for field inhomogeneities in spiral acquisitions, which will be further discussed in section 3.4.3.2.

3.4.2 Commercial and clinically available rt-MRI resources

Many studies have successfully imaged the upper vocal tract region with commercially available resources and on-the-fly reconstruction algorithms made available by the vendor. These have mostly focused on Cartesian acquisition real-time sequences derived from cardiac imaging and parallel imaging. Scott et. al investigated velopharyngeal closure at various spatial-temporal resolutions (10 to 20 fps, 1.9×1.9 mm^2 to 2.7×2.7 mm^2) using SSFP sequences and parallel imaging SENSE [4]. Vocal tract during singing and swallowing was imaged at 10 fps (1.7×2.7×6 mm^3) with a Cartesian sequence and parallel imaging GRAPPA [5, 43, 44]. Also, Martins et al. investigated velar movement in European Portuguese vowels at 14 fps (3.3×1.6×8 mm^3) with a similar Cartesian sequence and

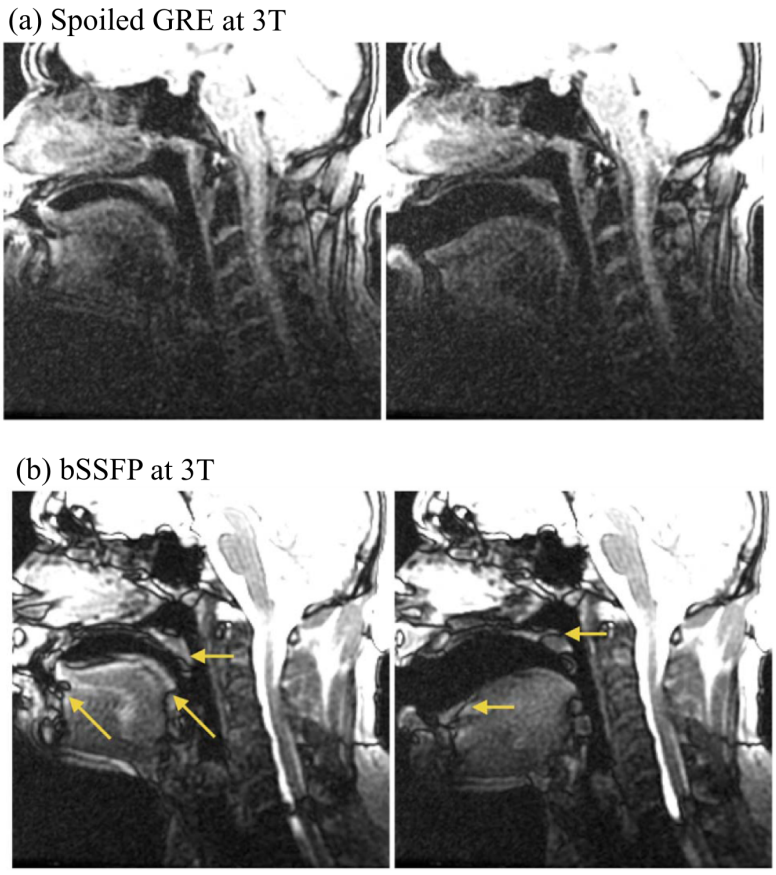


Figure 3.3: Example comparison of spoiled GRE *vs.* bSSFP rt-MRI of speech at 3 T. Although bSSFP sequences provide higher SNR performance, off-resonance artefacts are noticeable (particularly at higher field strengths), in this case characterised by signal void regions in the velopharyngeal area (yellow arrows). Image adapted from [39].

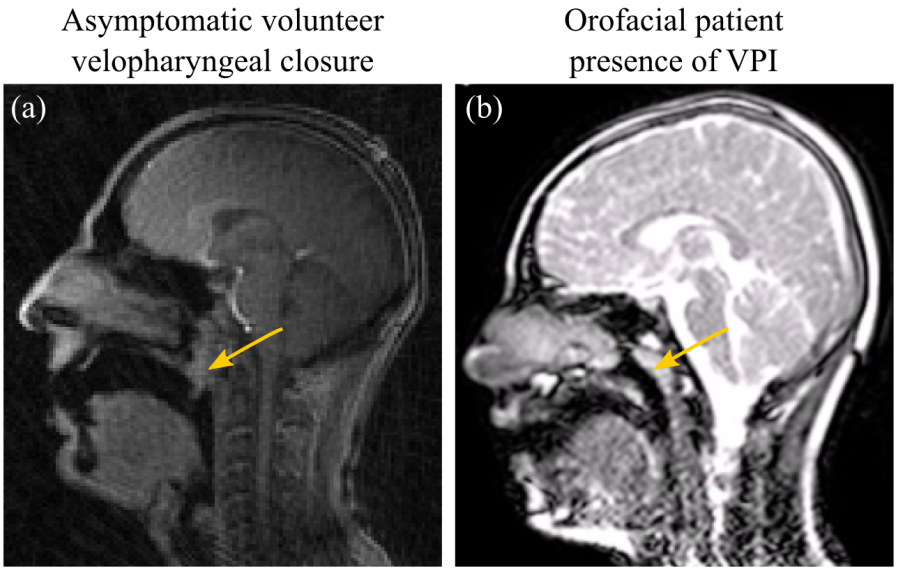


Figure 3.4: Example of rt-MRI in the assessment of velopharyngeal closure. Complete asymptomatic closure of the velopharyngeal port can be seen in an asymptomatic healthy volunteer (a) and incomplete closure, that is presence of VPI in an orofacial cleft patient (b).

GRAPPA set-up [42]. Another study by Kulinna-Cosentini et. al [45] compared the performance of TrueFISP *vs.* HASTE sequences at 3T in a cleft palate cohort, however limited to sub-optimal frame rates (∼2.75 fps).

Example of mid-sagittal images acquired with commercially available rt-MRI sequences, underlining differences between complete velopharyngeal closure on an assymptomatic volunteer and VPI present in an orofacial cleft patient can be seen in Figure 3.4.

3.4.3 Non-Cartesian rt-MRI

In addition to Cartesian widely available protocols, non-Cartesian acquisitions have been suggested to improve spatial-temporal resolution in speech rt-MRI. Intrinsic characteristics such as a more efficient coverage of k-space (e.g. spirals) and

oversampling of the centre region are advantageous to dynamic imaging. However, as noted in Chapter 2, reconstruction of non-Cartesian is intrinsically more complex than Cartesian k-space. The most commonly used non-Cartesian trajectories in speech are radial and spiral. As upper airway imaging is a fairly new field of interest, many rt-MRI sequences used in speech studies have been translated from cardiac applications. Thus, some reference to cardiac focused studies is given in the following sections.

3.4.3.1 Radial

As previously covered in Chapter 2, radial trajectories consist of sampling k-space in radial spokes that always cross $k_x, k_y = 0, 0$. Thus, each radial spoke always samples both low and high frequency information.

Previous dynamic studies [46, 47] resorting to radial encoding have focused on bSSFP sequences due to their superior SNR performance. Such studies were successful in performing free-breathing correction for cardiac dynamic acquisitions. A similar bSSFP radial sampling was also employed to perform real-time imaging of the heart. A maximum under-sampling acceleration of R = 10.7 compared to a full Cartesian data set (12 radial projections per frame) was achieved, corresponding to a temporal resolution of about 28 fps and a spatial resolution of 1.56×1.56 mm^2. Moderate artefacts were observed in the reconstructed images due to the high acceleration factor. However, a later study by Zhang et al. [48] suggested that high sensitivity of bSSFP sequences to off-resonance effects might compromise image quality of real-time MR studies at higher field strengths. Instead, a FLASH radial acquisition reconstructed with a view-sharing (sliding window) scheme was suggested. Real-time MRI of the heart was performed at 20 fps (2.0×2.0 mm^2) with negligible artefacts.

Further improvements in temporal resolution using the FLASH radial acquisition were later reported by the same group using a non-linear inversion reconstruction method [49] in cardiac rt-MRI studies [50, 51]. The technique was then

applied to visualise the upper vocal tract articulatory dynamics during natural speech [8, 52]. Imaging of the vocal tract was performed at 30 fps and a spatial resolution of 1.5×1.5 mm^2 with no major presence of susceptibility or motion artifacts.

More recently, and during the course of this PhD experimental work, Burdumy et al. [31] performed morphometric measurements of the vocal tract articulators using an accelerated golden-angle radial sequence and offline Conjugate Gradient SENSE (CG-SENSE) reconstruction pipeline (25 fps and 1.8×1.8 mm^2).

Due to the intrinsic over-sampling of k-space centre, radial acquisition is usually slower and less efficient in k-space coverage than for example spiral sampling. However, radial trajectories have shown an increased potential to withstand higher degrees of under-sampling and thus improve temporal resolution, while yielding adequate image quality.

3.4.3.2 Spiral

Similarly to radial, k-space spiral designs have been prescribed to improve spatial-temporal resolution of speech rt-MRI.

Over the years, researchers at the University of Southern California have focused on the application of spiral sampling to dynamic MRI. First applications to MR angiography at 1.5 T reported a raw temporal resolution of 6 fps, reconstructed up to 37 fps with a view-sharing method [53]. Due to the interleaved spiral acquisition, i.e. where each frame is acquired in multiple spiral interleaves, final display frame rate can be further improved by reconstructing each dynamic frame with the most recently acquired set of spiral interleaves. Although preliminary studies demonstrated promising use of spiral sampling for rt-MRI, the loss in SNR led to an increased interest to translate and test similar protocols at higher field strengths. Translation to 3 T reported similar spatial resolution (1.5×1.5 mm^2) and improved SNR [54]. Blood SNR improved by 57% when compared to the previous 1.5 T study, however at the compromise of temporal resolution

decrease to 24 fps.

A similar protocol was later implemented to assess speech and beatboxing performance [7, 55, 56]. A raw temporal resolution of 9 fps (1.8×1.8 mm^2) was reported, reconstructed up to 24 fps with view-sharing.

A study by Sutton et al. [57] compared an interleaved FLASH spiral sequence with view-sharing to a Cartesian protocol, while imaging the *levator veli palatini* (LVP) muscle. Spiral sampling allowed for comparable raw temporal resolution (6 fps) to the Cartesian protocol, however with improved final resolution up to 25 fps due to reconstruction with the view sharing method. The authors reported similar image quality for both protocols, suggesting that spiral protocols could provide similar image quality with improved temporal resolution.

Many studies have relied on view-sharing reconstruction to achieve higher display frame rates due to its simple implementation. However, and although reconstructed data presents dynamic frames closer together in time, no additional information is added to the raw data acquired. Thus, temporal fidelity is dependent on native acquired frame rate and temporal blurring may still be observed at higher frame rates if native resolution is inadequate. Improved raw temporal resolution (about 20 fps) was later reported by Bae et al. [6], using a reduced FOV of 120×120 mm^2. Saturation bands were applied every 6 spiral interleaves to eliminate signal outside the chosen FOV and avoid aliasing artefacts. A spatial resolution of 1.9×1.9 mm^2 and an improved view-shared frame rate of up to 30 fps was reported.

Alternatively, parallel imaging has also been frequently reported to improve spatial-temporal resolution with accelerated non-Cartesian data. Seiberlich et al. [58] imaged the heart at 55.5 fps while comparing the performance of radial and spiral sampling reconstructed with parallel imaging through-time GRAPPA. Since spiral trajectories intrinsically allow for a more efficient coverage of k-space, approximately 60% of time of an equivalent radial scan, further acceleration of data acquisitions was possible. The authors reported that for a specific spatial-temporal

resolution combination, spiral acquisitions required a much lower GRAPPA acceleration factor than radial acquisitions (e.g. R = 6 instead of R = 9 at 29 fps). Thus, spiral-acquired images presented less aliasing artefacts and improved SNR. Although a maximum temporal of 55.5 fps could be achieved, Seiberlich and colleagues suggested the best image quality *vs.* temporal resolution compromise to be 34 fps (1.95×1.95 mm^2) for cardiac applications. At this point, it is important to underline such a concept, where rt-MRI for clinical evaluation is not exclusively centred on obtaining the highest temporal acceleration possible, but instead in determining the best compromise possible for the clinical issue at hand.

More recently, and in parallel with the work presented in this thesis, Lingala et al. [59] also applied spiral GRAPPA to the real-time MRI of the upper airway, as will be further explored in experimental work in Chapter 6.

3.4.4 Multiple plane imaging of speech

A main advantage of MRI, when compared to conventional techniques such as x-ray videofluoroscopy, is its ability to acquire multiple imaging planes without repositioning the patient or limitation due to radiation dosage.

Some studies have used dynamic single-slice acquisition in multiple imaging views (e.g. sagittal, coronal and/or oblique), as opposed to a single mid-sagittal view. Sagar et al. imaged a cleft palate paediatric cohort with three imaging planes of the velopharyngeal port: sagittal for elevation and posterior movement of the velum, coronal to visualise movement of the lateral wall and oblique to evaluate sphincters performance [10]. Another study with a VPI cohort was limited to imaging velopharyngeal closure in two imaging planes, most commonly sagittal and oblique views [60]. However, both studies were performed at sub-optimal frame rates (∼2 fps).

Furthermore, multi-slice acquisition of interleaved imaging planes could provide with an accurate visualisation of the upper vocal tract during the production of the same speech task. Lingual articulation and vocal tract shaping in a healthy

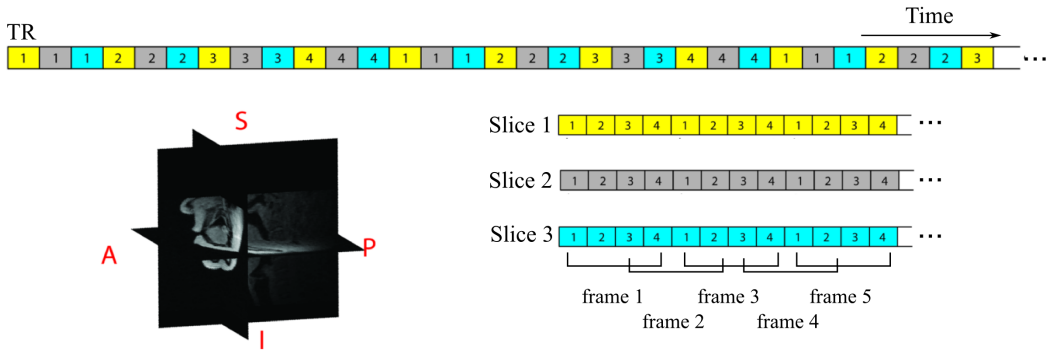


Figure 3.5: Schematic representation of interleaved multi-slice acquisition proposed by Kim et al. Slice 1 (yellow), slice 2 (gray) and slice 3 (blue) are acquired every TR with an interleaved spiral design. Example orientation of mid-sagittal, axial and coronal slices to image the upper vocal tract in-vivo are shown. Acquired multi-slice protocol is reconstructed using a view-sharing (sliding window) scheme where each frame is reconstructed with a frequent update of spiral interleaves. Image adapted from [61].

adult cohort was imaged using a 2-slice (sagittal and axial) and 3-slice (sagittal, coronal and axial) rt-MRI protocols [61, 62], implemented with an interleaved spiral acquisition and view-sharing reconstruction (schematic representation can be seen in Figure 3.5). The 2-slice acquisition had a spatial-temporal resolution of $2.4\times2.4\times6.0$ mm^3 and approximately 6.4fps (156 ms per frame), while the 3-slice protocol achieved a similar temporal resolution 6.1 fps (163 ms) in detriment of spatial resolution ($3.0\times3.0\times6$ mm^3). Kim et al. were able to demonstrate unique features such as the lateral motion of the pharynx and tongue grooving, visible in the axial slice, that were not seen in the mid-sagittal plane.

Recent work by Lingala et al. [63] demonstrated an improved temporal resolution of 36 ms (27.8 fps) for a 3-slice protocol ($2.5\times2.5\times5$ mm^3). This was achieved with a novel upper airway imaging protocol encompassing: a custom-built upper airway coil, golden-angle spiral acquisition and offline iterative constrained reconstruction.

Fu et al. [64] demonstrated promising results with data acquisition of 4 simul-

taneous slices at 25.5 fps and 8-slices at 12.8 fps ($2.2\times2.2\times6$ mm^3). Additionally, the same protocol was used to achieve 102 fps on a single-slice imaging plane. However, implementation was computationally challenging and inadequate for many clinical applications, taking up to 30 minutes to reconstruct a single slice of 10,000 frames at 102 fps.

Although multi-slice imaging has been shown to provide with additional insights into the mechanisms of human speech; spatial and temporal resolution must be sacrificed to the detriment of an increased number of interleaved slices. In addition, application of multi-slice protocols is yet to be translated to the clinical evaluation of speech impairments and having been mostly used for research purposes. Preliminary experiments in a VPI paediatric cohort with a single-slice protocol in a mid-sagittal and oblique plane (Figure 3.6) is currently under way by the research group where this thesis work was undertaken [65]. Accurate planning of the oblique plane in healthy subjects is fairly straightforward and based on angular measurements of the LVP muscle by Ettema et al. [66]. However, planning of the oblique plane in VPI patients with non-standard anatomical features is still challenging.

3.5 Summary

In this chapter, a summary of the production of speech and imaging of the upper vocal tract is given, with a particular focus on the use of rt-MRI. Although much work has been done in the field, speech MRI is still a fairly new area of research compared to others much more established, such as dynamic imaging of the heart. Furthermore, rt-MRI holds great potential for future translation into the clinical assessment of speech impairments. However, some constraints still remain to be explored. Much variability in acquisition protocol, reconstruction pipeline and imaging planes is still seen within the field [39]. The following experimental chapters in this thesis set out to address this and increase clinical reliability of

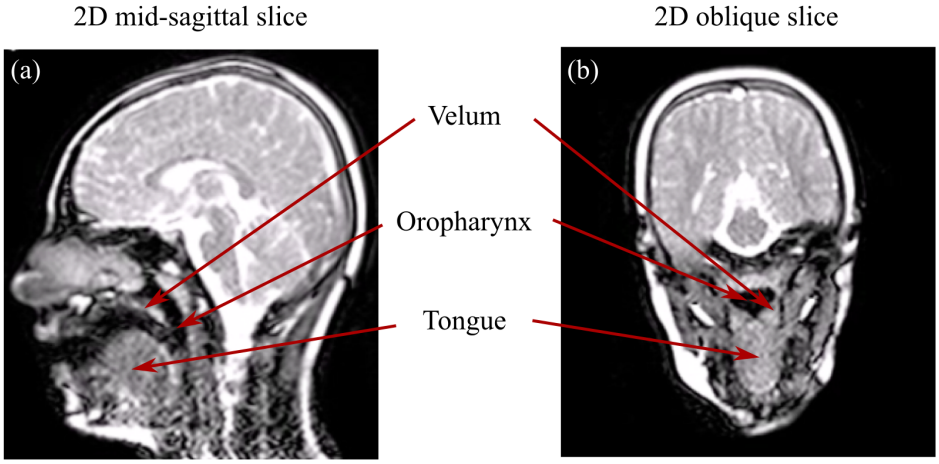


Figure 3.6: Example of real-time MRI of a VPI patient using commercially available protocols (EPI at 1.5 T, 1.9×1.9 mm^2) acquired using a 2D 10-mm thick slice in the mid-sagittal position and a 2D oblique slice.

speech rt-MRI. Two main steps can be identified: firstly, a comparison of imaging protocols for speech assessment is required in the field, secondly, further improvement with more advanced acquisition and reconstruction methodologies could bring new improvements to the field.

Chapter 4

Visualising and assessing velopharyngeal motion with commercial Cartesian and non-Cartesian rt-MRI

This chapter focuses on the use of commercial real-time MRI sequences, available as commercial products on current clinical scanners, to visualise and assess velopharyngeal motion and closure during speech. A quantitative comparison of commercial Cartesian and non-Cartesian imaging protocols at 1.5 T and 3 T, in regards to image quality and temporal resolution compromise, was performed. Optimal recommended protocols for the assessment of velopharyngeal motion are then suggested for both field strengths.

4.1 Introduction

As previously explored in section 3.4.2, many studies have imaged the upper vocal tract region during speech with commercial and widely available MR resources. These have mainly focused on Cartesian sequences and on-the-fly reconstruction

methods commercially offered by the manufacturer. Such studies have covered a wide range of speech tasks, both in healthy volunteers and patient cohorts [4, 5, 9, 42–44].

However, rt-MRI of the vocal tract is still an open field of research and much variability is seen in preferred acquisition protocols, reconstruction pipelines, number and/or orientation of imaging views. A recent review publication by Lingala et. al [39] provided an extensive set of recommended protocols and requirements for those entering the field. The authors also acknowledge a comparison of imaging protocols is challenging but would be most beneficial to the field of speech MRI. In addition, dynamic MRI is not currently used in the clinical assessment of speech disabilities and establishing reliable imaging protocols is essential. In summary, there is a clear need for a comparison of MR protocols in regards to image quality and temporal resolution compromise [39]. This could provide additional insights to researchers interested in the field as well as assisting with future translation to clinical assessment.

The main goal in this experimental chapter was to investigate the performance of commercial Cartesian and non-Cartesian rt-MRI sequences in speech assessment. No direct comparison of the trajectory sampling performance, Cartesian vs. non-Cartesian, was intended from these experiments as some differences in acquisition parameters are noticeable; instead it is intended to provide a comparison of recommended protocols for the assessment of velopharyngeal motion. At this stage, to ensure that resulting protocols can be easily reproduced across clinical scanners and research centres, only standard commercially available hardware, sequences and reconstruction pipelines were used. Investigation of imaging protocols for the visualisation of velopharyngeal motion was undertaken both at 1.5 T and 3 T, and optimal imaging protocols were then recommended for both field strengths.

4.2 Methodology: imaging velopharyngeal closure at 1.5 T and 3 T

4.2.1 Subjects and speech task

Five (2 males and 3 females, mean 42 years) and seven (2 females and 5 males, mean 32 years) adult subjects were recruited for the comparison of imaging protocols at 1.5 T and 3 T, respectively. None of the participants had any current diagnosed speech, language or hearing disabilities. Subjects were imaged in the supine position while performing a speech sample consisting of counting (1 to 10), non-sense nasal verbalisation (/za-na-za/, /zu-nu-zu/ and /zi-ni-zi/) and sustained phonation (/a/ as in arm and /i/ as in cheese). Participants were provided with and asked to repeat the selected speech sample before entering the MRI examination room.

Audio was simultaneously recorded using a fibre-optic MR-compatible microphone (FOMRI II, Optoacoustics, Or-Yehuda, Israel). The microphone end was strapped to the coil structure and placed adjacent and parallel to the lips of the subject. An adaptive noise-cancelling algorithm was used to reduce the background scanner noise [6]. Audio recording was started simultaneously with each MRI acquisition; however, subjects were instructed when to start phonation through the intercommunication system in order to allow for the noise cancellation algorithm to adjust. Audio recordings were later used for offline synchronisation with the real-time MR videos.

4.2.2 MRI data acquisition

Experimental work described in this chapter was performed using a 1.5 T Philips Achieva (Philips Healthcare, Best, the Netherlands) R3.2, a 3 T Philips Achieva Tx R3.3 scanners and a commercial 16-channel neurovascular head coil. A 2D mid-sagittal single-slice of the head and upper neck was acquired, with the shim

volume centred on the velum.

Preliminary experimental work was undertaken in order to identify suitable MR sequences, acquisition parameters and best compromise in image quality and temporal resolution. Preliminary optimisation work was performed with a phantom and in two volunteers for *in-vivo* testing. While Cartesian protocols were reproduced as described elsewhere [4], radial and spiral protocols had not been previously investigated and full protocol optimisation was undertaken. Contrary to previously reported optimal image quality when using Cartesian SSFP sequences [4], the use of SSFP sequences with radial and spiral trajectories was deemed inadequate for speech imaging during preliminary testing. Due to the presence of detrimental signal-nulling artefacts in the velopharyngeal region (Figure 4.1 - yellow arrows), a spoiled-GRE sequence (T1-FFE on the Philips system) was instead preferred for the non-Cartesian protocols.

Cartesian protocols at both field strengths were implemented with parallel imaging acceleration (SENSE) and a partial Fourier factor of 0.625. Non-Cartesian protocols at both field strengths were implemented with a gridding and view-sharing reconstruction available from the vendor. The view-sharing acceleration factor is defined as that per one fully acquired dynamic scan; a number of sub-dynamic scans defined by the acceleration factor are reconstructed by data-sharing. Therefore, temporal resolution is accelerated in proportion to the chosen view-sharing factor. Spiral protocols were implemented with 36 spiral interleaves and the shortest acquisition window possible (2 ms). Radial protocols were kept at 100% scan coverage, meaning that the number of radial spokes was matched to the size of acquisition matrix. All protocols at 1.5 T and 3 T were implemented with a single-slice thickness of 10 mm. As a quantitative comparison of non-Cartesian protocols against the established Cartesian protocols was intended, radial and spiral protocols were matched in spatial and temporal resolution to those of the Cartesian protocols. Two additional protocols were also implemented to investigate further temporal resolution improvement with non-Cartesian sequences. Full

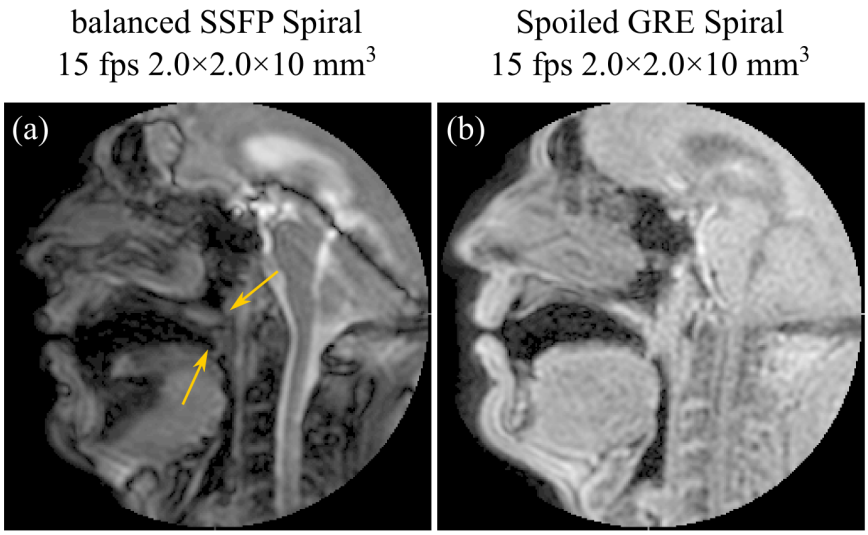


Figure 4.1: Example mid-sagittal images acquired during preliminary work to demonstrate differences in image quality for speech assessment between (a) balanced-SSFP spiral and (b) spoiled GRE spiral at otherwise similar acquisition parameters at 1.5 T. Due to the presence of signal-nulling artefacts in the velopharyngeal region (yellow arrows) that may hamper clinical assessment, a spoiled GRE non-Cartesian acquisition was preferred.

description of acquisition parameters for protocols implemented at 1.5 T and 3 T are shown in Table 4.1 and Table 4.2, respectively.

4.2.3 Data analysis

Images acquired with Cartesian and non-Cartesian protocols were analysed using both quantitative and qualitative measurements. All data sets were analysed using measurements of velum thickness and velum signal homogeneity, dynamic contrast-to-noise (CNR) of the velum and surrounding air cavity and qualitative visual scoring of image quality.

Spatial-temporal resolution	Sampling trajectory	TE/TR/α (ms/degrees)	FOV (mm)	View-sharing	SENSE
101 ms (10 fps) 1.9×1.9 mm^2	Cartesian	1.5/3.0/30	300×240	-	2.4
	Radial	2.3/5.1/10	180×180	5.0	-
	Spiral	1.0/5.1/10	190×190	2.0	-
65 ms (15 fps) 2.2×2.2 mm^2	Cartesian	1.4/2.8/30	300×240	-	3.0
	Radial	2.1/4.7/10	180×180	6.0	-
	Spiral	1.0/4.8/10	190×190	3.0	-
49 ms (20 fps) 2.7×2.7 mm^2	Cartesian	1.2/2.4/30	300×240	-	3.0
	Radial	1.9/4.1/10	180×180	6.0	-
	Spiral	1.0/4.8/10	190×190	4.0	-
45 ms (22 fps) 1.9×1.9 mm^2	Radial	2.3/5.0/10	170×170	9.0	-
	Spiral	1.0/5.1/10	190×190	4.0	-
38 ms (25 fps) 1.5×1.5 mm^2	Radial	2.7/5.9/10	170×170	16.0	-
	Spiral	1.0/6.3/10	190×190	6.0	-

Table 4.1: Acquisition parameters for protocol comparison at 1.5 T according to spatial-temporal resolution combination and sampling trajectory (Cartesian vs. radial vs. spiral).

Spatial-temporal resolution	Sampling trajectory	TE/TR/α (ms/degrees)	FOV (mm)	View-sharing	SENSE
105 ms (10 fps) 1.9×1.9 mm^2	Cartesian	1.0/2.2/15	300×240	-	1.8
	Radial	2.1/4.5/10	190×190	5.0	-
	Spiral	0.8/4.5/10	190×190	2.0	-
79 ms (14 fps) 2.2×2.2 mm^2	Cartesian	1.2/2.1/15	300×240	-	2.3
	Radial	1.9/4.1/10	190×190	6.0	-
	Spiral	0.9/4.4/10	190×190	3.0	-
48 ms (20 fps) 2.7×2.7 mm^2	Cartesian	0.9/2.5/15	300×240	-	2.4
	Radial	1.6/3.6/10	180×180	6.0	-
	Spiral	0.8/4.3/10	190×190	4.0	-
41 ms (22 fps) 1.9×1.9 mm^2	Radial	2.0/5.6/10	170×170	9.0	-
	Spiral	0.8/4.5/10	190×190	4.0	-
38 ms (25 fps) 1.5×1.5 mm^2	Radial	2.5/5.4/10	170×170	16.0	-
	Spiral	0.8/5.6/10	190×190	6.0	-

Table 4.2: Acquisition parameters for protocol comparison at 3 T according to spatial-temporal resolution combination and sampling trajectory (Cartesian vs. radial vs. spiral).

4.2.3.1 Velum thickness and signal homogeneity

Measurements of velum thickness and signal homogeneity were carried out using OsiriX 6.0.1 32bit (Pixmeo Sarl, Bernex, Switzerland). Both measurements were performed in two temporal positions in which the velum remains relatively motionless, from here onwards referred to as the elevated and relaxed velum positions. Velum measurements in the elevated position were performed in dynamic frames corresponding to the sustained phonation of /a/ and in the relaxed position, in dynamic frames during nasal breathing and no phonation. Velum thickness was defined as the measured distance between the velar knee and velar dimple [67]. Additionally, velum signal homogeneity was expressed as follows:

$$SH_{velum} = \frac{\overline{S_{velum}}}{\sigma_{velum}} \tag{4.1}$$

Where $\overline{S_{velum}}$ is the mean and σ_{velum} the standard deviation of the pixel intensities in the region-of-interest (ROI) selected in the velum region. This measurement gives an indication of the presence of artefacts or distortion in the selected area. As the assessment of speech disabilities with rt-MRI is yet to be translated into clinical practice, direct correlation between homogeneity of the velum's signal and clinical relevance still needs to be fully understood. However, low signal homogeneity of the velum could indicate a dynamic frame where the velum's boundaries are distorted or masked by artefacts, and thus, may hamper future clinical assessment of velopharyngeal motion.

4.2.3.2 Dynamic CNR velum/air

Dynamic CNR of the velum and surrounding air cavity were carried out using MATLAB (released 2014b, The MathWorks, Natick, MA). Intensity-time plots were obtained by selection of an intensity profile along the main direction of motion of the velum and stacking the profiles from adjacent time frames side-by-side (Figure 4.2). Intensity-time profiles of the velum allow for a representation of

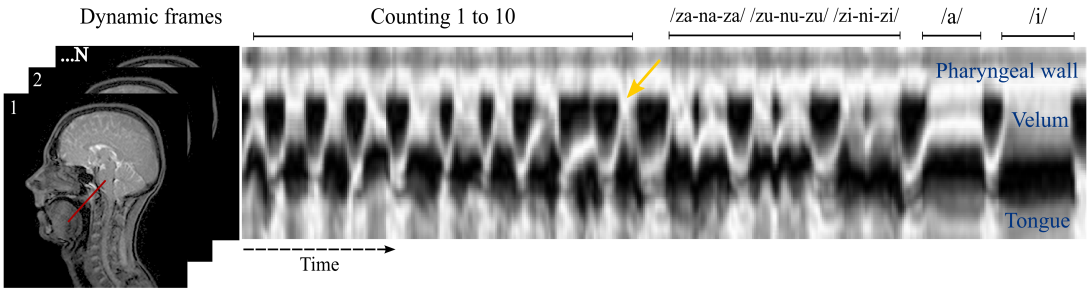


Figure 4.2: Schematic representation on obtaining intensity-time profiles from velopharyngeal rt-MRI data sets. Intensity-time profiles of velopharyngeal motion allow for a representation of velum motion throughout time. A velopharyngeal closure event is defined as each point of contact between the velum and the posterior pharyngeal wall (yellow arrow).

velum motion throughout acquisition, where the horizontal direction is representative of time. Dynamic CNR was measured in a short section of the intensity-time plots, considering two ROIs selected over the velum and neighbouring air cavity (refer to Figure 4.5):

$$CNR_{velum/air} = \frac{\overline{S_{velum}} - \overline{S_{aircavity}}}{\sigma_{aircavity}} \tag{4.2}$$

Where $\overline{S_{velum}}$ is the mean signal intensity in the ROI drawn over the velum, $\overline{S_{aircavity}}$ and $\sigma_{aircavity}$ are the mean and standard deviation of the signal intensity in the ROI drawn in the neighbouring air cavity.

4.2.3.3 Qualitative visual scoring

Image quality of MRI data sets was scored visually. Images were rated blindly and randomly using a five-point scale by two independent observers (one scoring done by the author and a second by another MRI physicist with previous speech MR experience). Observer 1 also scored the rt-MRI data sets a second time, approximately one month after the first scoring. Further details on the visual scoring scale are shown in Figure 4.3.

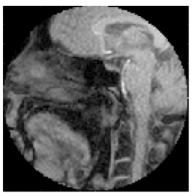
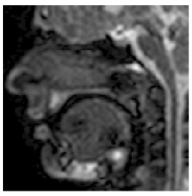
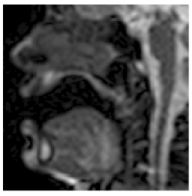
Scoring Scale	Example	Description of scoring scale level
(1) Non-diagnostic		• Velum can not be identified in most dynamic frames. • Image quality inadequate for closure assessment.
(2) Average		• Velum can be identified in most dynamic frames. • Closure assessment is severely hampered by distortion in the velopharyngeal region.
(3) Good		• Velum can be identified although with some interpolation through noise and/or artefacts. • Closure may be identified with some uncertainty.
(4) Very Good		• Velum is easily segmented and closure events identified with certainty. • Image noise and/or artefacts may be present in other regions of the image.
(5) Excellent		• Closure events can be identified with high certainty. • Image quality is overall good.

Figure 4.3: Scoring scale defined for visual assessment of the dynamic rt-MRI data sets. Grading was set on a five-point scoring scale from (1) non-diagnostic to (5) excellent. Corresponding image examples and description of scoring levels are given.

SPSS software (v.22, IBM, New York) was used to perform all statistical analysis. When comparing continuous variables of multiple measurements, e.g. the case of dynamic CNR comparison between different imaging protocols, repeated-measures one-way analysis of variance (ANOVA) was used, with a significance level of 0.05. A paired t-test with Bonferroni correction for multiple testing was used to identify significant pairs. Paired data sets, e.g. velum thickness at different velum positions, were compared using a two-tailed paired t-test. Image quality visual scoring was compared using a Kruskal-Wallis test. Significant pairs were identified with multiple pairwise comparisons of the Mann Whitney test and Bonferroni corrected significance level.

4.3 Results

4.3.1 Imaging velopharyngeal motion at 1.5 T

Example rt-MRI data sets of the selected speech task acquired with Cartesian, spiral and radial protocols at 10 fps and 20 fps at 1.5 T are shown in Figure 4.4.

Measurements of velum thickness in millimeters both at the elevated and relaxed velum positions are summarised in Table 4.3. Mean velum thickness across all data sets of the cohort ($N = 65$) was higher in the elevated position during the sustained phonation of /a/ (11.73 ± 1.77 mm) than in the velum relaxed position during nasal breathing (9.15 ± 1.51 mm) with significance ($p<0.0005$). No significant differences in velum thickness were found between protocols (vertical analysis, bottom row) measured in the relaxed and elevated velum positions. Additionally, no significant differences were found in velum thickness between sampling trajectories (Cartesian vs. radial vs. spiral) at both velum positions (horizontal analysis, far right column).

Signal homogeneity of the velum measured at both velum positions is summarised in Table 4.4. Measured signal homogeneity was greater in the relaxed posi-

tion (compared to the velum elevated position) across all protocols with Cartesian sampling ($p<0.05$). However, no significant differences in a pairwise comparison of signal homogeneity of the velum (between the relaxed and elevated positions) were found for most protocols of spiral sampling.

Considering a vertical analysis, no significant differences were found between imaging protocols (i.e. spatial-temporal resolution combinations) using Cartesian sampling (Table 4.4 bottom row). On the other hand, analysis underlined significant differences between imaging protocols for both velum positions for radial ($p<0.05$, Table 4.4 bottom row) and spiral ($p<0.01$ and $p<0.05$, Table 4.4 bottom row) acquisitions. Signal homogeneity in the relaxed position was found to be higher for protocol 10 fps when compared to protocol 25 fps ($p<0.01$). In the velum elevated position, significant pairs were found between signal homogeneity measured for protocols 10 fps and 20 fps ($p<0.05$) and protocols 20 fps and 22 fps ($p<0.05$) for radial acquisition, and between protocols 15 fps and 25 fps ($p<0.05$) for spiral acquisition.

On a horizontal analysis, some significant differences in signal homogeneity were identified when comparing sampling trajectories (Cartesian vs. radial vs. spiral) within equivalent spatial-temporal resolution protocols. Velum signal homogeneity was higher with spiral sampling compared to Cartesian at equivalent protocols (10, 15 and 20 fps) at both velar positions with significance (refer to Table 4.4). Comparison of spiral and radial acquisitions revealed significant differences in velum signal homogeneity for protocols 10 fps, 15 and 22 fps in the elevated velum position (refer to Table 4.4).

Examples of intensity-time displays of velopharyngeal motion acquired with spiral and radial protocols are shown in Figure 4.5. Overall, an adequate temporal depiction of velopharyngeal motion was observed across all spiral protocols and velopharyngeal closure events were easily identified. However, temporal displays of data sets acquired with radial protocols showed temporal blurring and distortion of the velum boundaries, rendering the identification of closure events inadequate.

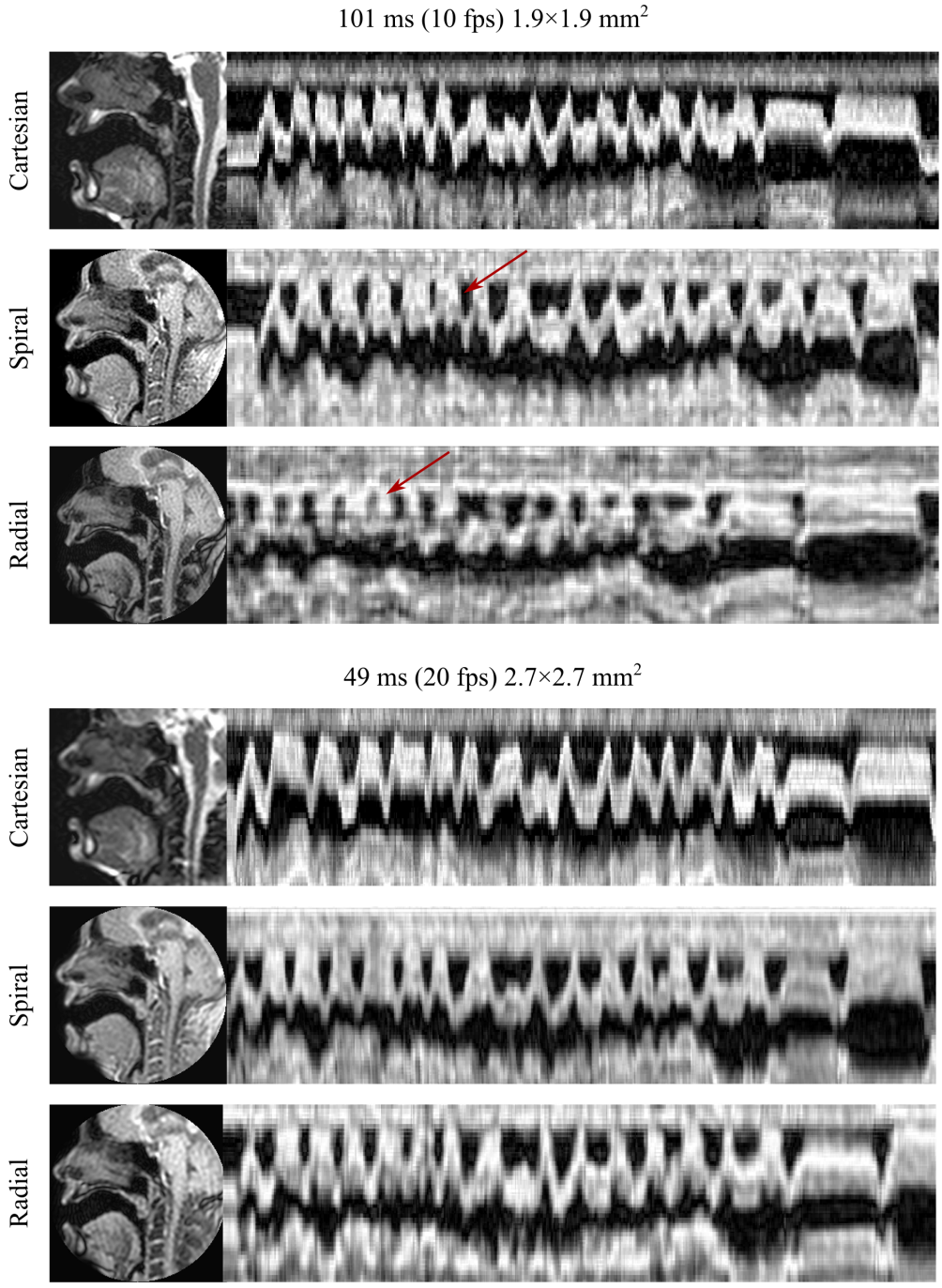


Figure 4.4: Mid-sagittal rt-MRI and intensity-time profiles during phonation at 1.5 T. Rt-MRI were acquired with Cartesian, spiral and radial protocols at 10 fps (1.9×1.9 mm^2) and 20 fps (2.7×2.7 mm^2). Spiral protocols demonstrated clear temporal depiction of velopharyngeal closure events while radial demonstrated increased temporal blurring (red arrows).

Protocols	**Cartesian**		t-test	**Spiral**		t-test	**Radial**		t-test	**p-value**
	Rel.	Elev.		Rel.	Elev.		Rel.	Elev.		
1	9.6 (1.4)	12.3 (2.1)	<0.1	8.7 (1.6)	11.7 (1.3)	<0.005	8.7 (2.2)	11.5 (2.0)	<0.5	0.65^{NS} / 0.77^{NS}
2	9.9 (1.6)	12.1 (2.6)	0.06	8.7 (1.4)	12.3 (0.8)	<0.01	9.1 (2.1)	12.0 (1.6)	<0.005	0.59^{NS} / 0.95^{NS}
3	10.0 (0.9)	12.2 (2.5)	0.08	9.3 (1.9)	12.6 (0.7)	<0.05	9.5 (2.0)	12.2 (2.1)	<0.05	0.75^{NS} / 0.92^{NS}
4	-	-	-	8.5 (1.2)	11.4 (1.3)	0.07	9.0 (1.4)	11.3 (1.3)	<0.05	0.56^{NS} / 0.86^{NS}
5	-	-	-	8.6 (1.0)	10.3 (1.8)	0.06	9.3 (1.3)	10.5 (2.1)	<0.01	0.39^{NS} / 0.83^{NS}
p-value	0.86^{NS}	0.99^{NS}	-	0.93^{NS}	0.06^{NS}	-	0.95^{NS}	0.66^{NS}	-	-

Table 4.3: Mean velum thickness and standard deviation in millimetres (mm) for all 1.5 T data sets (cohort of 5 subjects) measured in the relaxed (nasal breathing) and elevated (sustained phonation of /a/) velum positions. P-values refer to ANOVA analysis of velum thickness between protocols of the same sampling scheme (bottom row) and between sampling schemes (far right column, left values correspond to relaxed position comparison and right values to elevated position). NS-not significant.

Protocols	Cartesian		t-test	Spiral		t-test	Radial		t-test	p-value
	Rel.	Elev.		Rel.	Elev.		Rel.	Elev.		
1	3.71(0.51)^{a}	2.46(0.64)^{a}	<0.5	4.93(0.95)	4.17(0.84)^{c}	0.012	4.03(0.40)	2.94(0.27)	<0.01	<0.005/<0.0005
2	3.71(0.77)^{d}	2.20(0.42)^{a,c}	<0.5	4.63(1.07)	4.29(0.71)^{c}	0.56^{NS}	4.50(0.77)	3.27(0.27)	<0.05	<0.05/<0.0005
3	3.36(0.43)^{a,c}	2.29(0.41)^{a,b}	<0.05	4.59(0.77)	3.55(0.77)	0.15^{NS}	4.54(0.41)	3.76(0.55)	0.10^{NS}	<0.0005/<0.0005
4	-	-	-	4.17(0.40)	3.66(0.51)	0.18^{NS}	3.66(0.62)	2.84(0.38)	<0.05	0.07^{NS}/<0.01
5	-	-	-	3.75(0.52)	3.44(0.41)	0.30^{NS}	3.59(0.50)	2.98(0.47)	0.26^{NS}	0.59^{NS}/0.06^{NS}
p-value	0.14^{NS}	0.30^{NS}	-	<0.01	<0.05	-	<0.05	<0.05	-	-

Table 4.4: Mean and standard deviation velum signal homogeneity for all 1.5 T data sets (cohort of 5 subjects) measured from selected dynamic frames at the relaxed (nasal breathing) and elevated (sustained phonation of /a/) velum positions. P-values refer to ANOVA analysis of signal homogeneity between protocols of the same sampling scheme (bottom row) and between sampling trajectories within the same spatial-resolution compromise (far right column, left value corresponds to relaxed position and right value to elevated position). Post-hoc Bonferroni paired t-test was used to identify significant pairs, where: ^{a} p<0.0005 pairwise comparison to spiral acquisition, ^{b} p<0.0005 pairwise comparison to radial acquisition, ^{c} p<0.005 pairwise comparison to radial acquisition, ^{d} p<0.05 pairwise comparison to spiral. NS-not significant.

Particularly at higher frame rates, and with consequent loss in image quality, points of contact between the velum and the pharyngeal wall during speech are not clearly distinguished between consecutive velopharyngeal closures (red arrow), rendering the protocol non-diagnostic.

CNR of the velum and adjacent air cavity measured in the intensity-time displays are summarised in Table 4.5. No significant differences were found in CNR between 10 to 20 fps protocols (7.10 ± 1.87 to 6.54 ± 2.71) for Cartesian sampling. On the other hand, the same vertical analysis comparing protocols within the same sampling trajectory showed an increase in CNR between 10 fps and 20 fps protocols for radial (10.21 ± 1.74 vs. 13.27 ± 1.90, with borderline significance $p = 0.05$) and spiral (12.46 ± 1.31 vs. 17.68 ± 1.51, $p<0.0005$). At higher frames rates of 22 and 25 fps, a decrease in CNR was observed due to the choice of smaller pixel size. However, differences in CNR between the two fastest protocols for both radial and spiral trajectories were non-significant ($p = 0.99$). Additionally, comparison of 10 fps and 22 fps protocols for both non-Cartesian sampling methods (radial: 10.21 ± 1.74 vs. 7.37 ± 1.02, $p = 1.02$ and spiral: 12.46 ± 1.31 vs. 11.12 ± 0.59, $p = 0.83$) revealed no significant loss in CNR.

On a horizontal analysis, non-Cartesian acquisitions provided higher CNR than equivalent Cartesian protocols at 10, 15 and 20 fps. At higher frame rates (22 and 25 fps), spiral acquisitions demonstrated superior CNR than equivalent radial protocols (Table 4.5).

Results from qualitative visual scoring performed by two independent observers per sampling trajectory and per protocol are shown in Figure 4.6. Intra-observer agreement was good ($\kappa = 0.59$, $p<0.0005$) with scoring disagreement observed in 30.8% of the total cases and a maximum intra-observer difference of 1 scoring point. Inter-observer agreement was very good ($\kappa = 0.74$, $p<0.0005$) with scoring disagreement in 20% of the total data sets and also a maximum inter-observer difference of 1 score point. In 84.6% of the cases where observers did not agree, observer 2 consistently scored images higher by 1 scoring point than observer

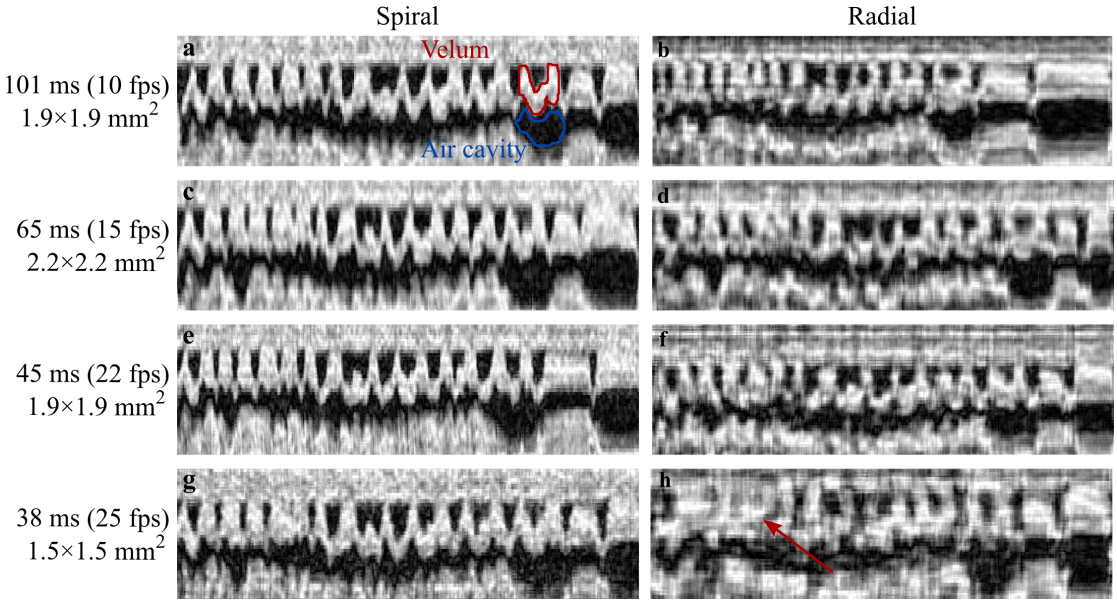


Figure 4.5: Example intensity-time displays of velum motion during speech for spiral (left) and radial (right) acquisitions at 101, 65, 45 and 38 ms per dynamic frame at 1.5 T. Selected ROIs in the velum (red) and the adjacent air cavity (blue) were used to calculate dynamic CNR. Temporal blurring and/or presence of velum distortions and artefacts can lead to misidentification of closure events (red arrow), rendering the protocol non-diagnostic.

1. In 18.5% of the total cases, corresponding to data sets acquired with spiral protocols at 10 fps and 25 fps, total scoring agreement between the two observers was observed.

Overall, spiral acquisitions provided superior qualitative scoring of image quality across protocols 10 to 20 fps than Cartesian protocols ($p < 0.01$). Although visual scoring of data acquired with spiral 10 fps was scored higher than the equivalent radial 10 fps (5.0 ± 0.0 vs. 2.9 ± 0.2, $p < 0.01$), no significant differences were found between spiral and radial samplings for protocols at 15 and 20 fps. At higher frame rates, qualitative scoring of image quality of spiral data sets was superior to that of radial protocols at 22 fps (4.4 ± 0.6 vs. 2.3 ± 0.7, $p < 0.01$) and at 25 fps (4.0 ± 0.0 vs. 1.6 ± 0.7, $p < 0.01$). In total, 12% of the analysed data sets were scored *5:Excellent*, all corresponding to spiral protocols.

Protocols	Cartesian	Spiral	Radial	p-value
1	7.10 (1.87)^{a,c}	12.46 (1.31)	10.21 (1.74)	< 0.005
2	6.67 (2.70) ^{b,d}	13.81 (1.23)	12.51 (1.92)	< 0.0005
3	6.54 (2.71)^{b,d}	17.68 (1.51)^{a}	13.27 (1.90)	< 0.0005
4	-	11.12 (0.59)	7.37 (1.02)	< 0.0005
5	-	9.89 (0.94)	6.82 (1.09)	< 0.005
p-value	0.93 ^{NS}	< 0.0005	< 0.0005	-

Table 4.5: Mean and standard deviation of CNR measured in a short section of the dynamic intensity-time displays. p-values refer to ANOVA analysis of CNR between sequences of the same sampling trajectory (vertical analysis, bottom row) and between sampling trajectories within the same spatial-temporal compromise (horizontal analysis, fair right column). Post-hoc Bonferroni paired t-test was used to identify statistically significant pairs, where: ^{a} p<0.05 pairwise comparison to radial ,^{b} p<0.005 pairwise comparison to radial, ^{c} p<0.005 pairwise comparison to spiral ,^{d} p<0.0005 pairwise comparison to spiral. NS-not significant.

4.3.2 Imaging velopharyngeal motion at 3 T

Examples of rt-MRI data sets of speech acquired with Cartesian, spiral and radial described protocols at 3 T at two spatial-temporal resolution combinations are shown in Figure 4.7.

Similarly to the results obtained at 1.5 T, average velum thickness was greater in the elevated velum position during the sustained phonation of /a/ (12.4 ± 1.0 mm) than in the velum relaxed position (10.3 ± 1.0 mm, p<0.0005). Measurements of velum thickness at the relaxed and elevated positions are shown in Table 4.6. Considering a vertical analysis, no significant differences were found across all imaging protocols (bottom row.) Also, no significant differences were found on a horizontal analysis, comparing different sampling trajectories within the same spatial-temporal resolution (far right column), with the exception of protocol 22

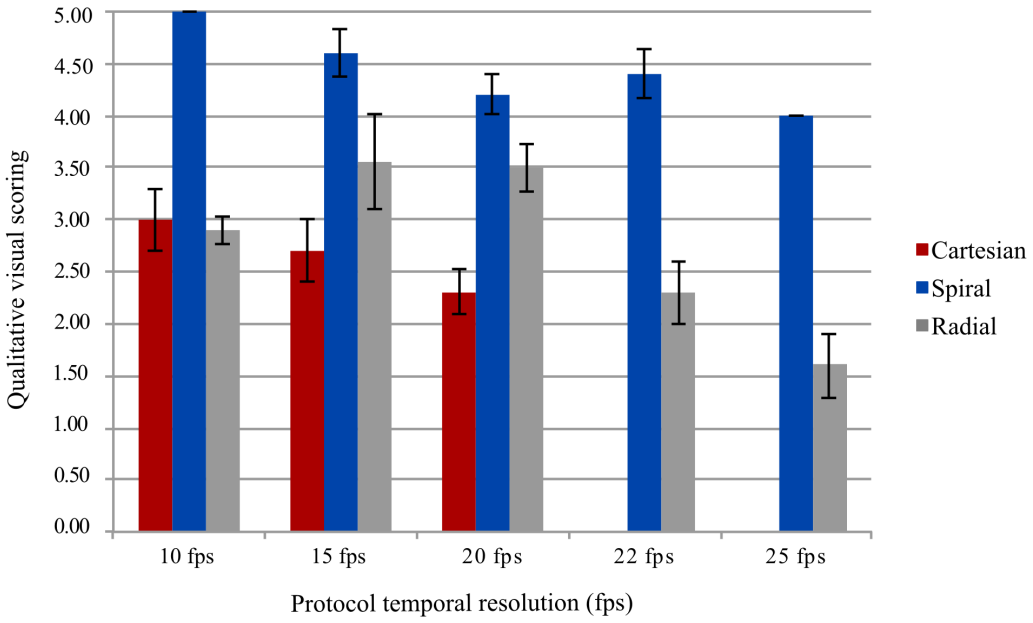


Figure 4.6: Qualitative visual scoring for 1.5 T data sets shown as mean (bar plot) and standard error (error bar) from two independent observers across different sampling trajectories (Cartesian, radial and spiral) and spatial-temporal resolution protocols (10, 15, 20, 22 and 25 fps).

fps: spiral vs. radial ($p<0.005$).

Signal homogeneity of the velum measured at both velum positions is summarised in Table 4.7. In accordance with previous speech imaging at 1.5 T, signal homogeneity within the velum boundaries was consistently greater for the relaxed position in pairwise comparison to the elevated position. However, significant differences were only determined between the two velum positions for all radial protocols and spiral protocols at 10 fps and 20 fps. Contrary to previous results at 1.5 T, no significant differences were found with paired t-test between relaxed and elevated positions for all Cartesian protocols.

On a vertical analysis, ANOVA analysis pointed out no significant differences between protocols for Cartesian, spiral and radial acquisitions (Table 4.7, bottom row). On the other hand, when comparing different trajectories within the same

spatial-temporal resolution protocol (horizontal analysis), some significant pairs were identified. Significant pairs were identified between Cartesian and radial trajectories at the elevated position for protocols 10 fps ($p < 0.005$) and 15 fps ($p < 0.01$) and between Cartesian and spiral at the elevated position for protocol 20 fps ($p < 0.05$).

Examples of intensity-time displays of velum motion acquired with non-Cartesian protocols are shown in Figure 4.8. Similarly to data sets at 1.5 T, good temporal depiction of velar movement can be observed across all spiral protocols and inadequate temporal display in data sets acquired with radial protocols.

Dynamic CNR measurements of the velum and surrounding air cavity are summarised in Table 4.8. No significant differences in CNR were found between data sets at 10, 15 and 20 fps for Cartesian, spiral and radial protocols ($p = 0.99$). Similarly, no significant differences in dynamic CNR were found between data sets at different spatial-temporal resolution protocols of radial sampling ($p = 0.06$), i.e. considering a vertical analysis. For spiral protocols, a decrease in CNR was noticeable when imaging at higher frame rates (26.57 ± 4.23 vs. 17.64 ± 3.99, $p = 0.001$), as expected with the choice of a smaller voxel size. However, differences in CNR between data sets acquired with sequences at 22 fps and 27 fps were non-significant (17.64 ± 3.99 vs. 19.09 ± 5.28, $p = 1.00$). Furthermore, spiral protocols provided superior CNR than Cartesian and radial protocols ($p < 0.0005$). At higher frame rates of 22 and 27 fps, spiral protocols also demonstrated superior CNR ($p < 0.005$) than radial protocols.

Results from qualitative visual scoring are shown in Figure 4.9. Kappa coefficient calculations demonstrated a good agreement both for intra-observer ($\kappa = 0.788$, 95% CI, 0.686 to 0.889, $p<0.0005$) and inter-observer ($\kappa = 0.786$, 95% confidence interval, 0.684 to 0.888, $p<0.0005$) ratings. Inter-observer scorings disagreed in 15.38% of the total data sets with a maximum difference of 1 scoring-point. Complete agreement between observers for data sets acquired with spiral 15 fps and radial 27 fps was observed. Overall, Cartesian data sets achieved superior

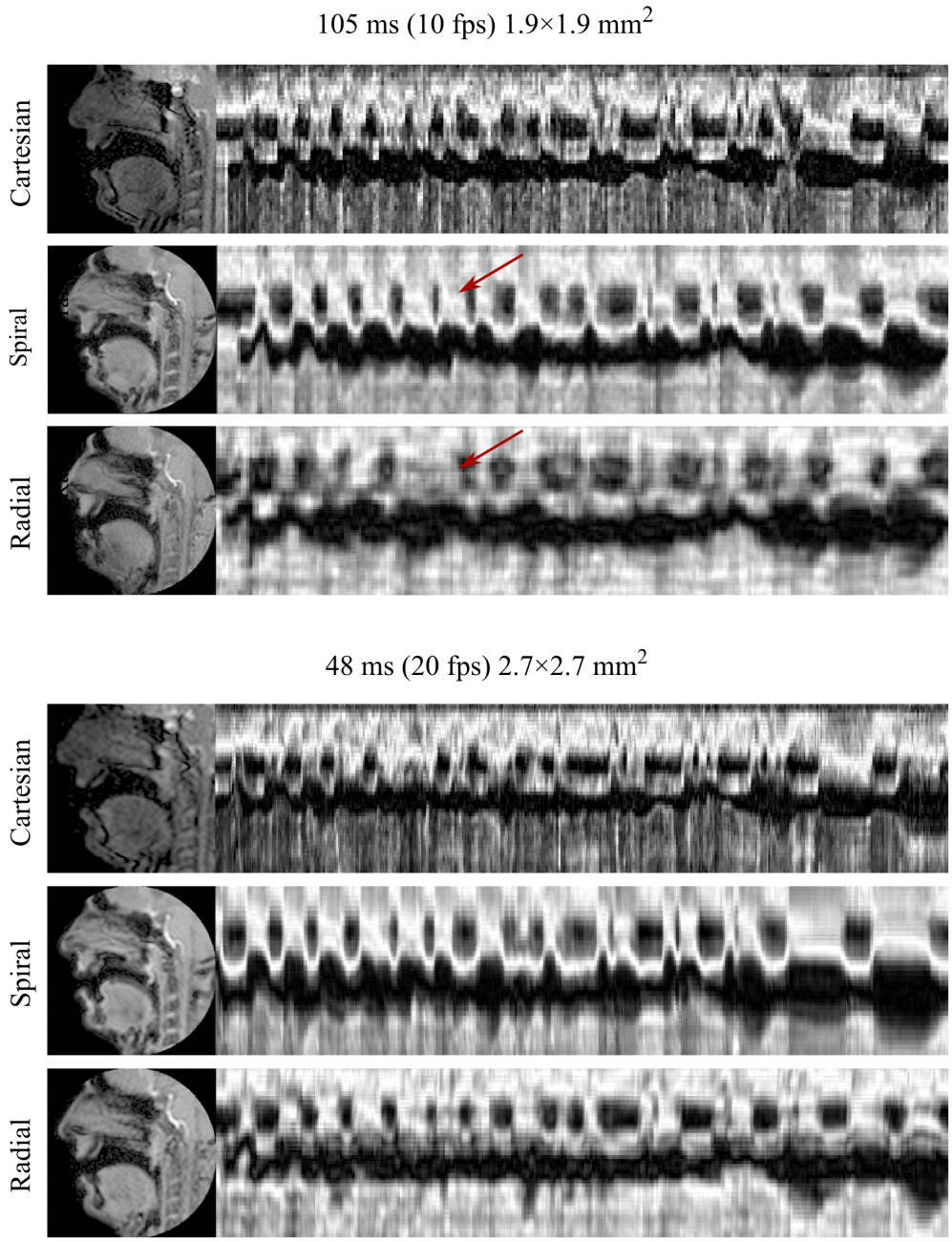


Figure 4.7: Dynamic mid-sagittal rt-MRI data sets and resulting intensity-time profiles during the phonation of selected speech task for one subject of the cohort at 3 T. Rt-MRI data sets were acquired with Cartesian, spiral and radial protocols at 10 fps (1.9×1.9 mm^2) and 20 fps (2.7×2.7 mm^2). Spiral protocols demonstrated clear temporal depiction of velopharyngeal closure events while radial protocols showed increased temporal blurring (red arrows).

Protocols	**Cartesian**		t-test	**Spiral**		t-test	**Radial**		t-test	**p-value**
	Rel.	Elev.		Rel.	Elev.		Rel.	Elev.		
1	10.3(0.9)	12.1(1.2)	<0.005	10.0(0.4)	12.3(0.8)	<0.0005	10.4(0.8)	12.8(1.0)	<0.0005	0.55^{NS}/0.49^{NS}
2	10.4(1.0)	12.2(1.1)	<0.05	10.4(1.2)	12.7(0.9)	<0.005	10.5(1.0)	12.0(1.1)	<0.005	0.47^{NS}/0.45^{NS}
3	10.4(1.0)	12.7(1.0)	<0.005	9.7(0.8)	13.0(1.4)	<0.0005	10.7(0.9)	12.4(1.0)	<0.0005	0.16^{NS}/0.69^{NS}
4	-	-	-	10.3(0.08)	13.2(1.0)	<0.005	9.9(1.1)	11.7(0.4)	<0.01	0.43^{NS}/<0.005
5	-	-	-	9.6(1.5)	12.3(0.9)	<0.01	10.3(1.3)	12.2(0.6)	<0.01	0.36^{NS}/0.89^{NS}
p-value	0.98^{NS}	0.66^{NS}	-	0.05^{NS}	0.37^{NS}	-	0.63^{NS}	0.19^{NS}	-	-

Table 4.6: Mean velum thickness and standard deviation in millimeters (mm) for all 3 T data sets (cohort of 7 subjects) measured in the relaxed (nasal breathing) and elevated (phonation of /a/) velum positions. P-values refer to ANOVA analysis of velum thickness between sequences of the same sampling scheme (bottom row) and between sampling schemes (far right column, left values correspond to relaxed position comparison and right values to elevated position).

Protocols	**Cartesian**		t-test	**Spiral**		t-test	**Radial**		t-test	**p-value**
	Rel.	Elev.		Rel.	Elev.		Rel.	Elev.		
1	3.95(0.99)	3.34(0.55)	0.17^{NS}	3.92(0.74)	2.71(0.32)	0.01	3.45(0.79)	2.32(0.56)	<0.05	0.48^{NS}/<0.005
2	3.65(0.87)	3.10(0.52)	0.12^{NS}	2.96(0.41)	2.64(0.41)	0.26^{NS}	3.65(0.89)	2.30(0.36)	<0.05	0.16/<0.05
3	3.54(0.77)	3.44(1.04)	0.82^{NS}	3.42(0.93)	2.26(0.42)	<0.01	3.58(0.83)	2.52(0.45)	<0.05	0.93/<0.05
4	-	-	-	3.53(0.85)	2.84(0.40)	0.08^{NS}	3.94(0.87)	2.31(0.46)	<0.005	0.40^{NS}/<0.05
5	-	-	-	3.85(1.22)	2.67(0.36)	0.08^{NS}	3.47(0.71)	2.25(0.44)	<0.005	0.48^{NS}/0.08^{NS}
p-value	0.69^{NS}	0.68^{NS}	-	0.09^{NS}	0.27^{NS}	-	0.84^{NS}	0.80^{NS}	-	-

Table 4.7: Mean and standard deviation velum signal homogeneity for all 3 T data sets (cohort of 7 subjects) measured from selected dynamic frames at the relaxed (nasal breathing) and elevated (sustained phonation of /a/) velum positions. P-values refer to ANOVA analysis of signal homogeneity between sequences of the same sampling scheme (bottom row) and between sampling trajectories within the same spatial-resolution compromise (far right column, left value corresponds to relaxed position and right value to elevated position).

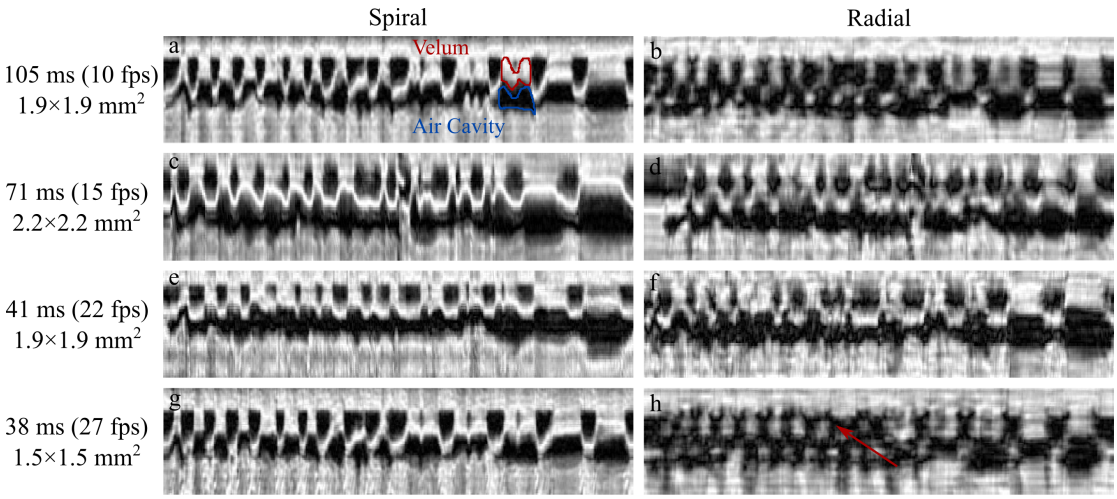


Figure 4.8: Example intensity-time displays of velum motion during speech for 3 T spiral (left) and radial (right) acquisitions at 105, 71, 41 and 38 ms per dynamic frame. Selected ROIs in the velum (red) and the neighbouring air cavity (blue) were used to calculate dynamic CNR. Temporal blurring and/or presence of velum distortion can lead to misidentification of closure events (red arrow).

visual scoring than spiral ($p<0.005$) and radial ($p<0.005$) across protocols 10 fps to 20 fps. Significant differences between visual scoring of spiral and radial data sets were mostly found at higher frame rates of 22 fps (2.93 ± 0.19 vs. 2.00 ± 0.58, $p<0.005$) and 27 fps (3.14 ± 0.38 vs. 1.00 ± 0.00, $p<0.0005$). No significant differences were found between visual scoring of protocols (10 vs. 15 vs. 20 fps) within Cartesian acquisitions.

4.4 Discussion

Several rt-MR imaging protocols were successfully implemented and compared for the visualisation and assessment of velopharyngeal motion during speech. Cartesian and non-Cartesian protocols were implemented using solely commercial hardware, sequences and reconstruction pipelines provided by the manufacturer to ensure being accessible to clinical and other non-research groups interested. Imaging protocols were considered for the two most commonly used field strengths in

Protocols	Cartesian	Spiral	Radial	p-value
1	8.82 (1.52)^{b}	25.94 (1.86) ^{a,c}	12.03 (2.92)	< 0.005
2	8.76 (1.91)	26.56 (3.83) ^{a,c}	11.56 (4.22)	< 0.0005
3	9.29 (3.32)	26.57 (4.23) ^{a,c}	13.63 (6.05)	< 0.0005
4	-	17.64 (3.99) ^{d}	8.90 (2.12)	< 0.0005
5	-	19.09 (5.28) ^{e}	8.11 (1.85)	< 0.005
p-value	0.90 ^{NS}	< 0.0005	0.06 ^{NS}	-

Table 4.8: Mean and standard deviation CNR for 3 T data sets measured in a short section of the intensity-time displays. P-values refer to ANOVA analysis of CNR measurements between the same sampling scheme and between spatial-temporal resolution sequence. Post-hoc Bonferroni paired t-test was used to identify significant pairs, where: ^{a} p<0.0005 comparison to Cartesian scheme, ^{b} p=0.05 comparison to radial scheme, ^{c} p<0.0005 comparison to radial scheme, ^{d} p<0.005 comparison to protocol 1 (10 fps) and ^{e} p<0.05 comparison to protocol 1 (10 fps).

the clinical environment, 1.5 T and 3 T. An optimised imaging protocol for the fast dynamic imaging of velopharyngeal closure has been proposed for both field strengths.

Velum thickness measured in the velum elevated position was greater than velum thickness in the relaxed position. This is in agreement with previous literature [4, 67], since the posterior-superior elevation of the velum results in the natural enlargement of the articulator in the sagittal plane. No significant differences in measured velum thickness were found when comparing protocols (i.e. spatial-temporal resolution combination) or when comparing Cartesian and non-Cartesian sampling at both field strengths. Such results seem to indicate that all imaging protocols consistently and accurately assess velum thickness, therefore not causing major distortions of its anatomical features in the mid-sagittal plane.

Dynamic CNR of the velum and surrounding air cavity was measured in a

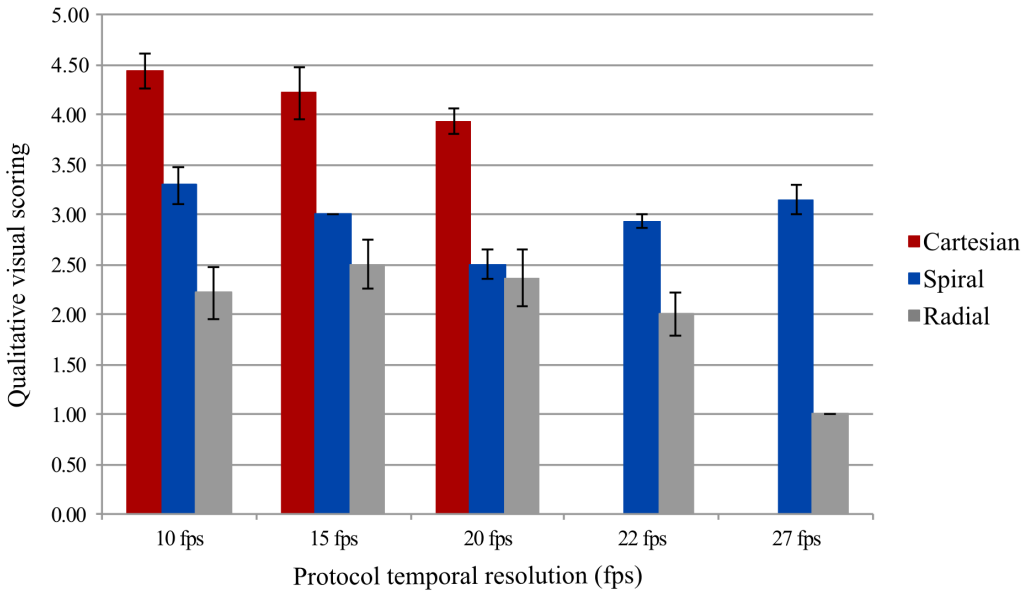


Figure 4.9: Qualitative visual scoring for 3 T data sets shown as mean (bar plot) and standard error (error bar) from two independent observers across different sampling trajectories (Cartesian, radial and spiral) and spatial-temporal resolution protocols (10, 15, 20, 22 and 27 fps).

short section of the dynamic intensity-time profiles. Overall and similarly for both field strengths, CNR of spiral data sets was greater across all protocols than CNR in Cartesian data sets. At higher frame rates, spiral protocols were optimal and provided superior CNR to equivalent radial protocols. As expected with the decrease in selected pixel size, CNR decreases for protocols at higher frame rates for both radial and spiral acquisitions. However, comparison between the two protocols revealed no significant difference in CNR, thus indicating that improvements in spatial-temporal resolution (from 22 fps, 1.9×1.9 mm^2 to 25 fps, 1.5×1.5 mm^2) could be achieved with no significant loss in CNR. For 1.5 T data sets, no significant difference in CNR was found between 10 fps and 22 fps protocols for both radial and spiral samplings. This result also indicates that both protocols allow to double temporal resolution, from 10 to 22 fps, while maintaining spatial

resolution (1.9×1.9 mm^2) with no significant loss in CNR performance.

At this point, intrinsic SNR characteristics of the sequences chosen for the Cartesian (SSFP) and non-Cartesian (RF spoiled GRE) protocols should also be discussed. Although SSFP sequences commonly allow for superior SNR compared to spoiled sequences [39], this particular measurement of CNR reflects both the intrinsic signal-to-noise ratio of the image and the presence of signal nulling areas/artefacts in the selected regions of interest, i.e. velum and adjacent oral cavity. Therefore, although spiral acquisitions were implemented with a spoiled-GRE sequence (Philips T1-FFE), due to the reduced sensitivity to off-resonance and overall improvement in signal homogeneity of the velum region, measured CNR was superior than of SSFP Cartesian data sets. Since a pairwise comparison between all imaging protocols was intended, and not a true quantitative measure of image SNR, CNR of the velum and adjacent cavity more accurately reflects how well the velum can be distinguished from the surrounding background.

While quantitative measures of image quality (e.g. velum CNR) may provide valuable information, due to the future translation of such protocols to clinical practice, a qualitative visual assessment of image quality is equally important. Visual scoring of the rt-MRI videos provided additional insights into the comparison of diagnostic value of the imaging protocols, particularly when comparing different field strengths.

When imaging at 1.5 T, spirals provided superior qualitative scoring of image quality across protocols 10 fps to 20 fps than equivalent Cartesian data sets. For 22 fps and 25 fps protocols, spiral acquisitions also demonstrated optimal visual image quality for velopharyngeal assessment and superior to equivalent radial data sets. Spiral 10 fps protocols were consistently scored as *5 - Excellent* across all subjects and observers. However, obtaining optimal image quality at 3 T was increasingly challenging and hampered by the larger effect of off-resonance at the velum-air boundaries. When imaging at 3 T in the current study, some inter-subject variability was found in the imaging protocol that scored highest in

the visual assessment. Results suggested that when imaging at 3 T, Cartesian protocols are on average superior in visual image quality and should therefore be preferred as an overall recommended protocol to image velopharyngeal motion. However, and because inter-subject variability increases with field strength, some particular cases may benefit from being imaged with the suggested spiral protocol. These could be cases such as imaging at a higher temporal resolution while maintaining spatial resolution (e.g. spiral 22 fps, 1.9×1.9 mm^2), need for higher CNR performance (e.g. improved segmentation of vocal tract shape) or superior image quality in some particular subjects.

Therefore, in order to make recommendations on imaging protocols for velopharyngeal imaging, I recommend a two-step approach. Firstly, a visual assessment of image quality should be considered if future clinical translation is intended. Secondly, when visual image quality is above a certain threshold and/or when more than one imaging protocol are of a similar image quality, it is worth taking into account additional quantitative criteria such as CNR performance, improvement in sharpness of dynamic intensity-time profiles or distortions of velum thickness. Also, if further post-processing (e.g. segmentation) of data sets is intended as the main goal of a study, giving priority to CNR performance and sharpness of velum boundaries may be preferred to a slightly higher visual scored data set with inferior CNR (e.g. spiral 10 fps vs. Cartesian 10 fps).

With this clinical set-up, radial protocols were overall inadequate for the fast dynamic imaging of the velum. As previously covered, reconstruction of non-Cartesian data was performed with vendor provided gridding and view-sharing scheme. While this method allows for a higher display frame rate, no additional data is added beyond the native rate of acquisition. To fulfil the Nyquist criterion at the edges of k-space, radial sampling oversamples the centre and takes longer to fully acquire one dynamic frame than spiral counterparts [12]. Therefore, radial protocols in this study were less optimal since the native frame rate was intrinsically lower and a higher view-sharing interpolation was required (Table 4.1) to

achieve an equivalent frame rate, thus increasing temporal blurring.

As a final note, the true potential of non-Cartesian imaging relies on the undersampling of k-space to further accelerate data acquisition, and thus improve frame rate. Image quality could then be recovered with more complex and non-standard reconstruction methods, that further improve the performance or rt-MRI in speech assessment. Therefore, some of those methods are further investigated in the following chapters of this thesis.

4.5 Summary

In summary, adequate visualisation of velopharyngeal motion during natural speech with commercial and widely available MR resources was possible at 1.5 T and 3 T. Both Cartesian and spiral imaging protocols demonstrated adequate image quality overall and temporal resolution compromise for the speech application at focus. However, obtaining adequate image quality at 3T is more challenging and subject-dependent than at 1.5 T. Based on the present results, it would be advisable to prefer imaging at 1.5 T when possible, using the recommended protocols at around 10 to 15 fps to ensure adequate temporal depiction of velopharyngeal motion [41]. Although Cartesian protocols provide adequate depiction of vocal articulators and can be used if the only option available, higher frame rates can only be obtained by Cartesian protocols at the compromise of spatial resolution (e.g. 20 fps at 2.7×2.7 mm^2) and CNR. On the other hand, spiral protocols provided superior CNR and visual image quality than otherwise equivalent Cartesian protocols and are thus recommended at 1.5 T imaging if available. When imaging at 3 T, obtaining adequate image quality might be increasingly challenging and hampered by the larger effect of off-resonance at the velum/air boundaries. Therefore, pre-testing both Cartesian and spiral protocols on a shorter speech sample might be advisable before deciding on a final protocol for each subject.

Results presented in this experimental Chapter have been published in the following peer-reviewed publications:

A C Freitas, M Wylezinska, M J Birch, S E Petersen, M E Miquel. Comparison of Cartesian and non-Cartesian real-time MRI sequences at 1.5T to assess velar motion and velopharyngeal closure during speech. *PLoS ONE* 11(4), 2016.

A C Freitas, M Ruthven, R Boubertakh, M E Miquel. Real-time speech MRI: commercial Cartesian and non-Cartesian sequences at 3T and feasibility of offline TGV reconstruction to visualise velopharyngeal motion. *European Journal of Medical Physics*, 2018. (*in press*)

Chapter 5

Feasibility of accelerated golden-angle radial imaging with TGV-regularised reconstruction in velopharyngeal imaging

This chapter sets out to investigate the application of novel suggested iterative SENSE-reconstruction methodology with a Total Generalized Variation (TGV) regularisation for the reconstruction of accelerated non-Cartesian data in the visualisation and assessment of velopharyngeal motion.

5.1 Introduction

As previously discussed (Chapter 2), acquisition of MRI data is much slower than other imaging modalities (e.g. x-ray video fluoroscopy in speech studies). Temporal resolution of MRI can be greatly improved by reducing the number of acquired phase-encoding lines (or in the radial case, radial projections). However, reducing the number of phase-encoding lines (or increasing Δ_{kr} as described in Section 2.3.4) causes aliasing in the final reconstructed image, which translates as

a periodic replication of the imaged object onto itself, for a Cartesian k-space, or streaking artefacts, in a radial non-Cartesian k-space. Hence, accelerating image acquisition comes at the cost of decreased image quality (lower SNR) and presence of aliasing artefacts.

Some reconstruction methods have suggested the use of iterative regularised techniques to minimise aliasing artefacts in the final reconstructed image. These methods have gained particular attention in dynamic MRI due to their adequate performance in minimising overall artefacts level as well as not being reliant on fully-sampled training data [39].

5.1.1 Reconstruction as an inverse problem

The process of MR reconstruction is an inverse problem. From the MR signal equation 2.13, a matrix form can be described as follows:

$$y = Mx \tag{5.1}$$

where x is a vectorised form of the unknown final image, M is the encoding matrix and y the vector of all the measured MR signals. The size of y is given by $n_c \times n_k$, respectively the number of coil elements and the number of k-space sample points.

The previously described encoding matrix M can be defined as a combination of Fourier terms at k-space values k and each coil specific sensitivity weight (C_l) at a specific image pixel p (R_p) [26]:

$$M = e^{ikR_p} C_l(R_p) \tag{5.2}$$

The final reconstructed image x can be calculated by the direct inversion of equation 5.1, therefore given as:

$$x = M^{\dagger} y \tag{5.3}$$

where $M^{\dagger}$ describes the Moore-Penrose pseudo-inverse of the encoding matrix:

$$M^{\dagger} = (M^H M)^{-1} M^H \tag{5.4}$$

Where M^H is the conjugate transpose of the encoding matrix M. However, calculation of the direct inverse of M would be computationally prohibitive in current MRI reconstructions due to its size. For example, for a radial k-space with 64 radial spokes on a 128×128 matrix and 8-channel receiver coil, M would have the size of $(8{\times}128{\times}64){\times}128^2$ and $\sim 128^6$ operations would be necessary to calculate $M^{\dagger}$ [26].

An alternative and more efficient way is to solve the problem iteratively, i.e. an image vector x is iteratively estimated until "best matching" the measured data y. How well the image estimate x fits y is given by the calculation of the squared difference (or L_2 norm) of the two variables:

$$\phi(x) = \|Mx - y\|_2^2 \tag{5.5}$$

Therefore, the best image vector x is said to be the one that minimises the cost function $\phi(x)$, given as:

$$x \Rightarrow minimises(\phi(x)) \tag{5.6}$$

A commonly used method to solve equation 5.6 is the Conjugate Gradient (CG) method. CG was firstly introduced in 1952 [68] and since then has been frequently used in MRI reconstruction problems [69].

The iterative search of the CG scheme can also be interpreted by the selection of a direction of search for the estimated final reconstructed image x. This is done by calculating the derivative of the cost function ($\phi(x)$) [70], here defined as the gradient $\bigtriangledown(\phi(x))$:

$$\bigtriangledown(\phi(x)) = M^H Mx - M^H y = M^H(Mx - y) \tag{5.7}$$

Thus, the most adequate estimation of x is usually found when the derivative is reduced to zero. By observing equation 5.7, the calculation steps of the CG algorithm can also be understood. In each iteration, the estimated image x is transformed to the spatial frequency domain by multiplication with the encoding matrix M. Then, the difference $(Mx - y)$ (also known as *residuum*) is calculated in order to estimate how well the estimate of x fits the measured data y. Finally, the gradient is calculated by mapping back the residuum into image space by multiplication with M^H [71].

5.1.2 Regularisation

Going back to an image reconstructed from the optimisation of the cost function in equation 5.5, one issue remains. Because the cost function only assesses the accuracy of the image estimation at the measured k-space locations (in this case, at the location of the radial projections), inaccuracy may still be present between those k-space locations [71] and the reconstructed image may still present a high degree of aliasing artefacts. In order to improve image quality, an additional term (or terms) known as regularisation or constraint terms ($R_i(x)$) can be introduced to pre-condition the cost function. Thus, the regularised cost function becomes:

$$\phi(x) = \|Mx - y\|_2^2 + \sum \lambda R_i(x) \tag{5.8}$$

Where lambda λ is a user-selected factor that weights the tendency of the estimated x to be more strongly matched to the measured data y (data fidelity) or to the pre-conditioning performed by the regularisation terms (*a priori* knowledge). Many regularisation terms have been prescribed in MR reconstruction problems, such as the Tikhonov [72], Total Variation (TV) [73] and Total Generalized Variation (TGV) [74].

5.1.2.1 Total Generalized Variation (TGV) regularisation

The TV regularisation was initially introduced by Rudin et al. [73] and recently reported in the reconstruction of under-sampled radial MR data by Block et al. [70]. The TV penalty is based on the assumption that the true object being imaged consists of areas of constant intensity. The TV penalty can be described by the operator that calculates the derivatives in the x- (D_x) and y- directions (D_y):

$$R_{TV}(x) = \sum \mid D_x(x) \mid + \mid D_y(x) \mid \tag{5.9}$$

Then, the first derivative at the pixel location (m, n) is given by:

$$D_x(m, n) = x(m, n) - x(m - 1, n) \tag{5.10}$$

$$D_y(m, n) = x(m, n) - x(m, n - 1) \tag{5.11}$$

Due to the assumption of fairly constant areas of image intensity, excessive regularisation of the TV operator (i.e. λ too large) may lead to final reconstructed images that look "blocky" and artificial in aspect. However, if carefully selecting the regularisation weight, particular cases such as the imaging of vocal tract structures (with good contrast at the tissue-air boundaries) may greatly benefit from TV regularisation, as recently suggested by Burdumy et al. [31, 75].

While TV regularisation has been frequently used in the denoising of MR images, the assumption of constant intensity might not always be applicable to MR (e.g. due to intrinsic tissue signal variations within and between organs and local B_1 inhomogeneities). Thus, a second order derivative TGV penalty has been recently described.

TGV regularisation was firstly introduced by Bredies et al. in 2010 [76] as a mathematical concept for image denoising, later applied to medical imaging and MRI applications [74]. Recent work by Knoll et al. demonstrated that TGV regularisation could be applied to both MR image denoising (e.g. high-resolution

brain images) as well as reconstruction of under-sampled non-Cartesian MR images. The TGV penalty can be defined as a minimisation problem in itself, given as follows:

$$R_{TGV} = min\alpha_1 \int |\bigtriangledown x - v|dx + \alpha_0 \int |\varepsilon(v)|dx \tag{5.12}$$

where v refers to the complex field vectors and $\varepsilon(v) = \frac{1}{2}(\bigtriangledown v - \bigtriangledown v^T)$ denotes the symmetrised derivative. The ratio of positive weights given by α_1 and α_0 allows to balance between the first and second derivative. This ratio is given by $\alpha_1 = \frac{\alpha_0}{2}$.

5.1.3 Radial golden-angle (GA) scheme

As previously explored, final display temporal resolution can be further increased with the use of a temporal view-sharing scheme (also known as sliding window). Here, radial projections acquired in one dynamic acquisition can be shared with the temporally adjacent frames allowing for the reconstruction of intermediate temporal frames, thus increasing final display frame rate. However, if a conventional radial angle increment is used (refer to equation 2.24) and only a portion of acquired radial profiles are selected for view-sharing (i.e. reconstruction of intermediate temporal frames), the distribution of sampled data in k-space is non-uniform.

On the other hand, a golden-angle (GA) scheme samples radial projections (or spiral interleaves) uniformly around k-space since the rotation angle between consecutive projections is constant. In fact, this uniform sampling is maintained even when the number of radial profiles decreases (under-sampling of k-space) as the GA rotation angle does not depend on the total number of acquired profiles (refer to equation 2.25). This characteristic is of great interest, in particular for dynamic imaging, since a flexible retrospective selection of temporal resolution is possible. By selecting adjacent radial profiles that were acquired in a golden-

angle fashion (uniform coverage of k-space), flexibility is being added to data reconstruction in regards to the selection of final display temporal resolution. Secondly, the golden-angle radial sampling of each dynamic frame is in its essence unique and thus resulting aliasing artefacts in image space are highly incoherent in the time dimension [77].

5.1.4 Application of GA acquisition and constrained reconstruction to speech imaging

A radial GA acquisition scheme with an iterative CG-SENSE reconstruction has been recently described by Burdumy et al. [31] to perform morphometric measurements of the vocal tract. Here, the reconstruction step was changed from 25 radial spokes (originally required to reconstruct one dynamic frame) to 18, thus allowing sharing profiles between dynamic frames. This allowed for a final temporal resolution of about 25 fps. This work was developed and published during the experimental work undertaken in this thesis.

In addition, the use of golden-angle sampling is not restricted to radial trajectories and it has been successfully applied to other non-Cartesian trajectories (e.g. spiral or stack-of-radials). Kim et al. [78] demonstrated a first application of golden-angle spiral interleaved scheme for improved depiction of rapid tongue tip motion. A similar application of golden-angle to spiral sampling was later demonstrated by Lingala et al. using both a GRAPPA [59] and TV-regularised iterative [63] reconstruction, both works were also published during the experimental work described in this thesis.

5.2 Methodology

5.2.1 MRI data acquisition

Experimental work described in this chapter was undertaken on a Philips Achieva 3 T Tx MRI scanner (Philips Healthcare, Best, the Netherlands), software release R3.3 and a 16-channel neurovascular coil by the same vendor. Four adult subjects (4 males, mean age 37 years) were recruited for this experimental work. Subjects were imaged in the supine position while performing the previously described speech sample (Chapter 4).

A 2D real-time mid-sagittal slice of the head and neck was acquired with a spoiled-GRE radial (Philips: T1-FFE) trajectory. This was implemented with the following imaging parameters: FOV = 240×240 mm^2, base matrix = 108×108, TE/TR = 1.84/4.04 ms, flip angle (α) = 10 degrees, 2D single-slice 10 mm thick with acquired in-plane spatial resolution = 2.2×2.2 mm^2. In a first experimental work, aimed at assessing the feasibility of a TGV-SENSE method in velopharyngeal imaging, work was undertaken using Philips commercial 2D radial trajectory. In this case, radial projections followed a readout-alternation sampling (i.e. each consecutive readout projection is acquired in alternating directions), with a rotation angle given as $\triangle\phi = \frac{\pi}{N_{spokes}}$, where $0 \leq \phi \leq \pi$. Considering an azimuthal rotation within $0 \leq \phi \leq \pi$, 108 radial spokes correspond to the fully sampled equivalent Cartesian k-space and thus defined as R = 1. In addition, radial data sets under-sampled in the angular direction were acquired at acceleration factors of R = 2.5, 4.9 and 5.7, corresponding to 43 (177 ms, ∼6 fps), 22 (89 ms, ∼11.2 fps) and 19 (61 ms, ∼17 fps) radial spokes.

5.2.2 Radial golden-angle sampling

In addition to the standard radial trajectory described in section 5.2.1, a golden-angle sampling was also implemented. Sequence pulse programming of the re-

quired acquisition sequence was undertaken using available Philips Pulse Programming Environment (PPE) R 3.3. Firstly, a user defined variable named ACQ USER ANGLE was defined, accepting either active YES or deactivate NO. This ACQ USER ANGLE option was added to the scanning parameters menu of the MRI scanner, if set to NO the commercial standard radial trajectory was still available for acquisition. On the other hand, if ACQ USER ANGLE = YES, then a second user-defined variable appeared named ACQ GOLDEN ANGLE. This variable was set to receive an integer number between 0 and 360 (corresponding to degrees). Using this second field in the menu, the user was then able to define the chosen angle step between consecutive radial profiles. If the radial sampling is performed between $0 \leq \phi \leq \pi$, then $\phi_{GA} \approx 111.25$ degrees, if the radial spokes were acquired over $0 \leq \phi \leq 2\pi$ then $\phi_{GA} \approx 2 \times 111.25$ degrees. For all the experimental work described in this chapter, a $0 \leq \phi \leq \pi$ sampling scheme was considered. Implementation of these acquisition parameters (radial sampling scheme and angle rotation) were set to allow the greatest flexibility for future research work, without any need to rewrite the pulse sequence source code. Since readout alternation is mainly advantageous in consecutive radial projections that are acquired closer together in k-space (in order to reduce eddy currents from gradient switching), readout alternation of the radial projections was deactivated when a golden-angle mode was selected.

5.2.3 Image reconstruction with TGV-regularisation

Under-sampled radial k-space was reconstructed with an iterative TGV regularisation method, described here and made available by Knoll et al. [74]. During scanning, image data was reconstructed with scanner vendor available gridding in order to provide on-the-fly guidance in scan planning. Raw data was then exported from the scanner system and read from raw format into a MATLAB matrix as $n_{readout} \times n_{spokes} \times n_{coil} \times n_{dynamics}$. Offline iterative TGV-regularised reconstructions were implemented within MATLAB (The MathWorks Inc., Natick, MA,

R2014 b) on a 2.5 GHz Intel Core i7, 16 Gb computer.

As described in the original publication [74], coil-by-coil images were estimated via a standard L_2 norm reconstruction (no TGV regularisation) and convolution with a smoothing kernel. Coil sensitivity maps for each coil element were obtained by dividing each low-resolution image of the coil element by the combined sum-of-squares image (SOS). Estimation of coil sensitivity maps was stopped at 50 iterations as no further visual improvement was noticeable.

Regularised TGV-SENSE reconstruction was implemented according to the original formulation described in section 5.1.2.1 and a NUFFT function (open-source toolbox [24]). A zero image was used to initialise the CG iterative reconstruction.

Selection of the TGV regularisation parameter was mainly selected by visual inspection of the rt-MRI data sets, looking for the optimal balance between reduction of aliasing artefacts and good identification of the vocal tract shape. For this purpose, data sets corresponding to two subjects of the cohort were reconstructed at different weights of TGV regularisation.

5.2.4 Data analysis

Measurements of CNR at the velum/air cavity boundary during phonation of the sound /zi-ni-zi/ were undertaken (same implementation as in Chapter 4). Sharpness of the intensity-time profiles was measured in the temporal (along multiple time dynamic frames) and spatial (along the same intensity-profile) directions. Measurement of sharpness in both directions was defined as the inverse of the distance between the 20-80% intensity points along each selected profile. A total of four intensity profiles at different locations of the speech task were selected for each measurement (temporal direction and spatial direction) and averaged.

Statistical analysis was undertaken using SPSS software (v. 22, IBM, New York). For continuous variables of multiple measurements, such as comparing CNR between reconstructed images, a repeated-measures one-way analysis of vari-

ance (ANOVA) analysis was performed, with a significance level of 0.05. Multiple Bonferroni paired t-test was used to identify significant pairs.

5.3 Results

5.3.1 Feasibility of accelerated radial rt-MRI with TGV-regularised reconstruction

Speech rt-MRI data were successfully acquired and exported for offline TGV reconstruction in all subjects of the cohort. An example of coil sensitivity estimation is shown in Figure 5.1. Firstly, coil-by-coil images were estimated as described in Section 5.2.3, from the under-sampled raw data. An inner loop of 50 iterations was used in the calculation of the coil sensitivity maps, as previously described. As expected from the iterative estimation of the coil sensitivities from under-sampled radial k-space data, aliasing artefacts are noticeable in the background of each coil map image.

Selection of the TGV-regularisation weight (α) was based on visual assessment of both the rt-MRI mid-sagittal dynamic frames and resulting intensity-time profiles. An example of the process of visual selection of the regularisation weight (α) in two subjects of the cohort is shown in Figure 5.2. When α is set to zero, no TGV regularisation is applied and the final reconstructed image is solely based on the original cost function, described in equation 5.5, thus resulting in a lower image quality with higher levels of residual noise and streaking artefacts. On the other hand, overweighting of the regularisation term (such as the moderate case of $\alpha = 0.5$ and the extreme case of $\alpha = 1$), might remove key anatomical features in the reconstructed image due to the effect of over-smoothing of the regularisation term. In this case, an optimal regularisation of $\alpha = 0.1$ was selected for velopharyngeal imaging, presenting a good compromise between removal of streaking artefacts and adequate depiction of the vocal tract shape and temporal display of

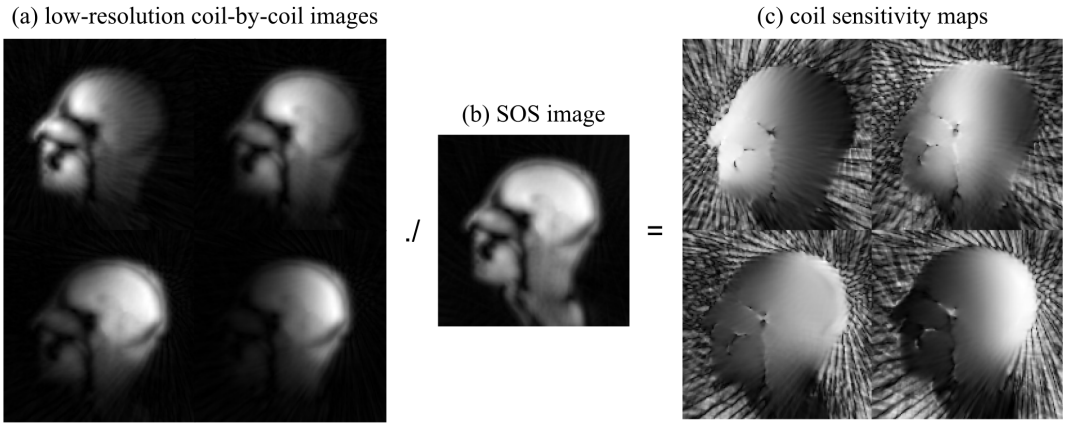


Figure 5.1: (a) Low-resolution coil-by-coil images calculated from raw under-sampled radial data (R = 2.5) corresponding to 4-elements selected from a 16-channel neurovascular coil and L_2 norm iterative reconstruction. (b) Sum-of-squares (SOS) image calculated from the 16-coil elements reconstructed. (c) Example of estimated coil sensitivity maps corresponding to the same 4 coil elements, resulting from dividing low-resolution coil-by-coil images (a) by the SOS image (b).

the intensity-time profiles. All data sets were reconstructed using 200 iterations as no further visual improvement on image quality was noticeable past it.

Measurements of CNR, temporal and spatial sharpness at the velum/air cavity boundary of two subjects of the cohort for different regularisation weights are summarised in Table 5.1. When no regularisation is applied ($\alpha = 0$), an increased level of noise can be seen in the reconstructed images (Figure 5.2) and a decreased CNR performance of 10.01 $\pm$ 2.16. However, for the weighting of $\alpha = 0.1$, here selected for velopharyngeal imaging, a superior CNR of 21.15 $\pm$ 2.31 ($p<0.0005$, compared to no-regularisation) and temporal sharpness 0.18 $\pm$ 0.07 were measured. If over-weighting is applied ($\alpha = 0.5$ and $\alpha = 1$), a decrease in CNR performance was observed. Thus, underlying the choice of $\alpha = 0.1$ as the optimal TGV weighting for velopharyngeal imaging in this particular set-up. Although a small increase was noticeable in temporal and spatial sharpness comparing recommended $\alpha = 0.1$ and other α values, no significant difference was found with

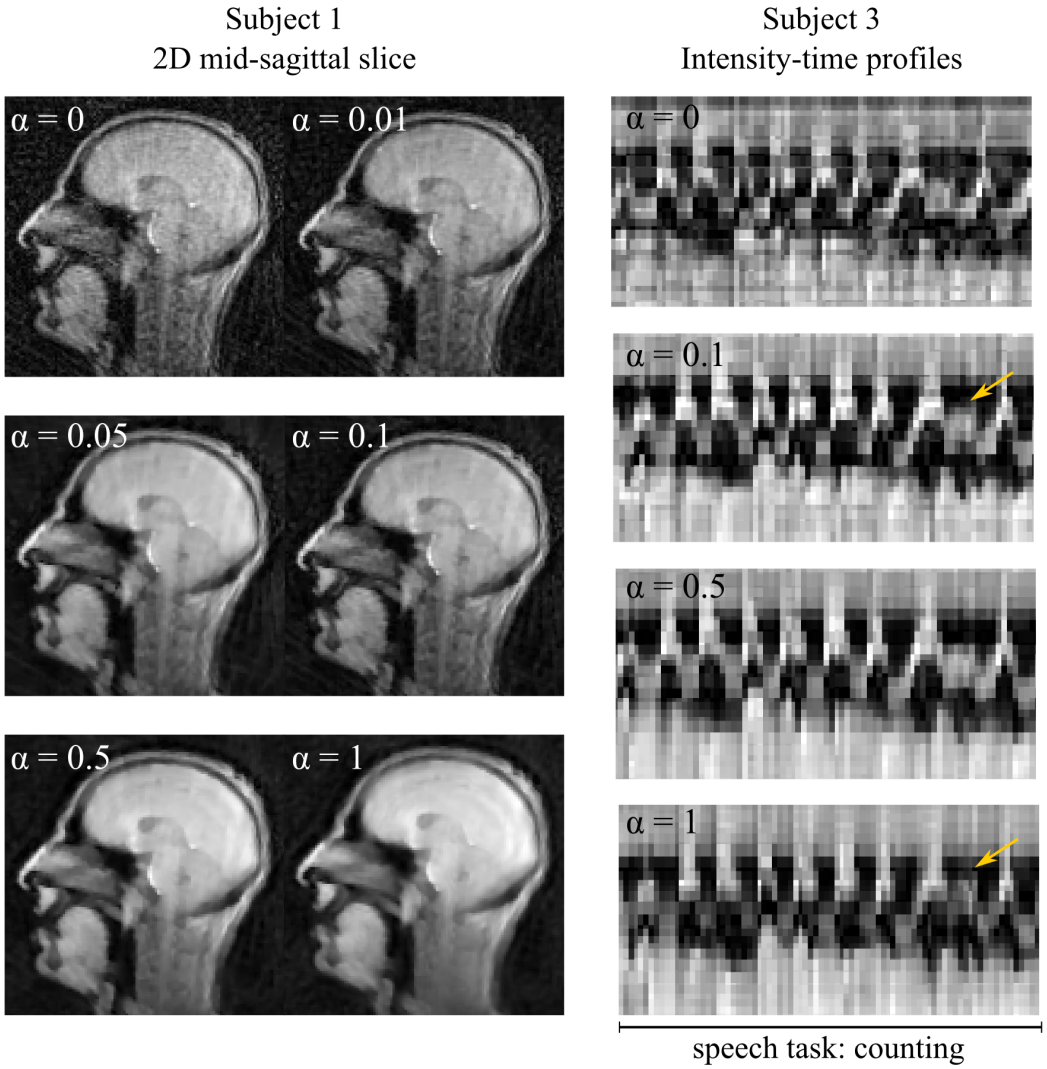


Figure 5.2: Effect of TGV regularisation in iterative reconstruction of accelerated radial (R = 2.5, 43 radial projections) rt-MRI of speech in two subjects of the cohort. No regularisation ($\alpha = 0$) is equivalent to a L_2 norm cost function formulation, thus presenting a higher presence of aliasing artefacts (streaking). Increase of the TGV regularisation allows to decrease the presence of aliasing artefacts and random noise. An optimal weighting of $\alpha = 0.1$ was selected for velopharyngeal imaging. Over-weighting of the TGV operator may lead to exaggerated smoothing and blurring of some velopharyngeal closure events (yellow arrows).

ANOVA analysis (Table 5.1 bottom row).

Accelerated radial data sets acquired during speech task performance were reconstructed with the TGV-regularisation method ($\alpha = 0.1$) and conventional NUFFT for comparison. Examples of 2D mid-sagittal dynamic slices of the head during the sustained phonation of the sound /a/ are shown in Figure 5.3. Data sets were acquired with fully-sampled radial imaging (R = 1, 108 projections) and accelerated to R = 2.5, 4.9 and 5.7, corresponding to 43, 22 and 19 radial projections. Radial data sets (Figure 5.3 a) acquired with 108 radial projections correspond to the fully-sampled radial k-space compared to a Cartesian sampling (R = 1), hence the near absence of aliasing artefacts. However, for the accelerated radial data sets, since the Nyquist criterion is not fulfilled and sampling of k-space data is insufficient, final reconstructed images present increased level of aliasing artefacts (streaking). As expected, the degree of aliasing (and consequent increase in visual presence of streaking artefacts) increases in relation to the degree of radial acceleration (i.e. less radial projections sampled) for data reconstructed with conventional NUFFT. On the other hand, some improvement in minimising the degree of aliasing artefacts was observed with TGV regularisation, particularly at higher degrees of acceleration.

5.3.2 Assessing velopharyngeal motion with accelerated radial TGV

Figure 5.4 compares intensity-time displays of velopharyngeal motion from fully-sampled radial data (R = 1, 108 radial projections), accelerated radial R = 2.5 (43 projections) and R = 4.9 (22 radial projections) reconstructed with the selected TGV regularisation ($\alpha = 0.1$). In the case of R = 1, due to the reduced temporal resolution (454 ms per dynamic frame), accurate sampling of the velopharyngeal motion and closure pattern is not possible (yellow arrows). However, acceleration of radial data by a factor of 2.5 (43 projections) allowed for a native temporal

TGV weighting	CNR Velum/Air	Temporal Sharpness	Spatial Sharpness
$\alpha = 0$	10.01 (2.16)	0.11 (0.03)	0.11 (0.09)
$\alpha = 0.1$	21.15 (2.31)^{a,b}	0.18 (0.07)	0.12 (0.05)
$\alpha = 0.5$	17.39 (3.34)	0.15(0.04)	0.08 (0.02)
$\alpha = 1$	14.98 (1.20)	0.14 (0.09)	0.06 (0.01)
p-value	< 0.0005	0.28^{NS}	0.26^{NS}

Table 5.1: Mean and standard deviation measurements of CNR in the velum/air boundary, velum/air sharpness in the temporal and spatial directions performed in two subjects of the cohort to evaluate the effect of TGV regularisation in reconstructed images. p-value refers to one-way ANOVA analysis. ^{a} p<0.0005 pairwise comparison to $\alpha = 0$ and ^{b} p<0.05 pairwise comparison to $\alpha = 0.5$ and $\alpha = 1$. NS-not significant.

resolution of 170 ms ($\sim$ 7 fps), adequately sampling the movement of the velum at a normal speech rate (17). Further improvements in identifying consecutive velum closure events was noticeable with highly accelerated radial data (R = 4.9, $\sim$11 fps). Identification of closure events corresponding to the rapid motion of the velum during the sound /za-na-za/ was clearer in intensity-time displays corresponding to data sets at $\sim$11 fps compared to $\sim$7 fps (red arrows). However, increase in temporal resolution is hampered by decreased image quality in the dynamic data sets (mid-sagittal frames).

Measurements of CNR, temporal sharpness and spatial sharpness at the velum/air cavity boundaries were performed in a short section of the intensity-time profiles of the entire cohort. Results are summarised in Table 5.2. Average CNR of the velum and surrounding air cavity with no regularisation was 9.82 $\pm$ 3.36; and 0.11 $\pm$ 0.03 and 0.07 $\pm$ 0.03 of temporal and spatial sharpness. Data sets reconstructed with the selected regularisation ($\alpha = 0.1$) showed an increase in measured CNR to 14.42 $\pm$ 2.13 (p<0.05). No significant differences were found in the measure-

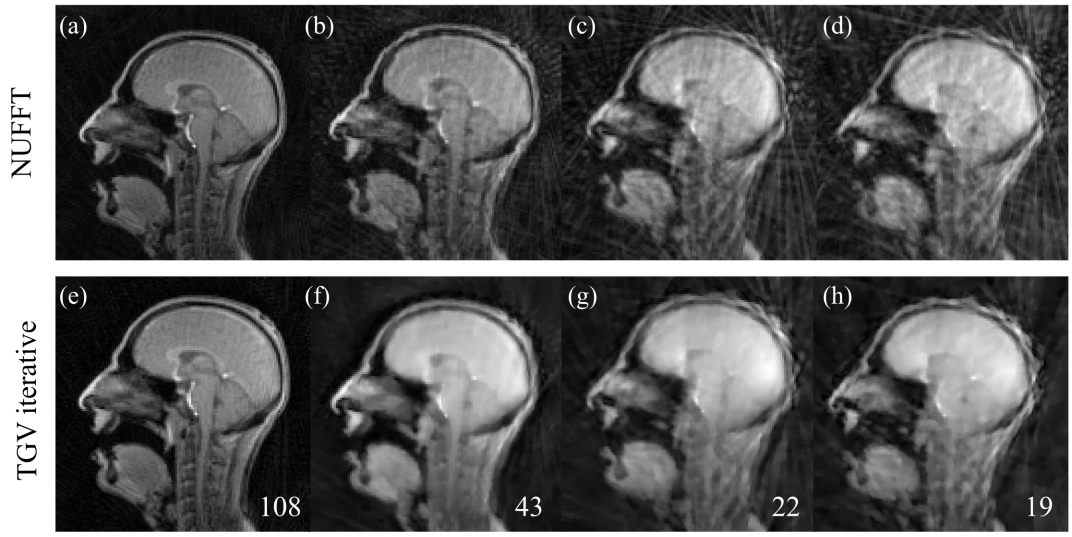


Figure 5.3: Example of dynamic mid-sagittal frames from rt-MRI data sets during the phonation of the sound /a/ acquired with 108, 43, 22 and 19 radial projections, corresponding to acceleration factors R = 1, 2.5, 4.9 and 5.7. Data sets were reconstructed using NUFFT reconstruction (top) and TGV regularised reconstruction $\alpha = 0.1$ (bottom). Bottom right numbers indicate the number of radial projections acquired in each data set.

ments of temporal and spatial sharpness between data sets reconstructed with no regularisation and the selected $\alpha = 0.1$.

An example of rt-MRI of speech with accelerated radial TGV reconstruction during the phonation of the sound /zu-nu-zu/ is shown in Figure 5.5. A total of 14 consecutive dynamic frames at a frame rate of 177 ms are shown. Although some aliasing is present in the final reconstructed images due to the under-sampling of k-space (2.5-fold), motion of the vocal articulators during /zu-nu-zu/ can be observed. Pronunciation of the sound /zu.../ results in the elevation of the velum until full velopharyngeal closure, followed by the rapid lowering of the velum for the nasal sound /...nu.../. Lastly, velopharyngeal closure occurs again for the last /...zu/ phonation with elevation of the tip of the tongue (yellow arrow).

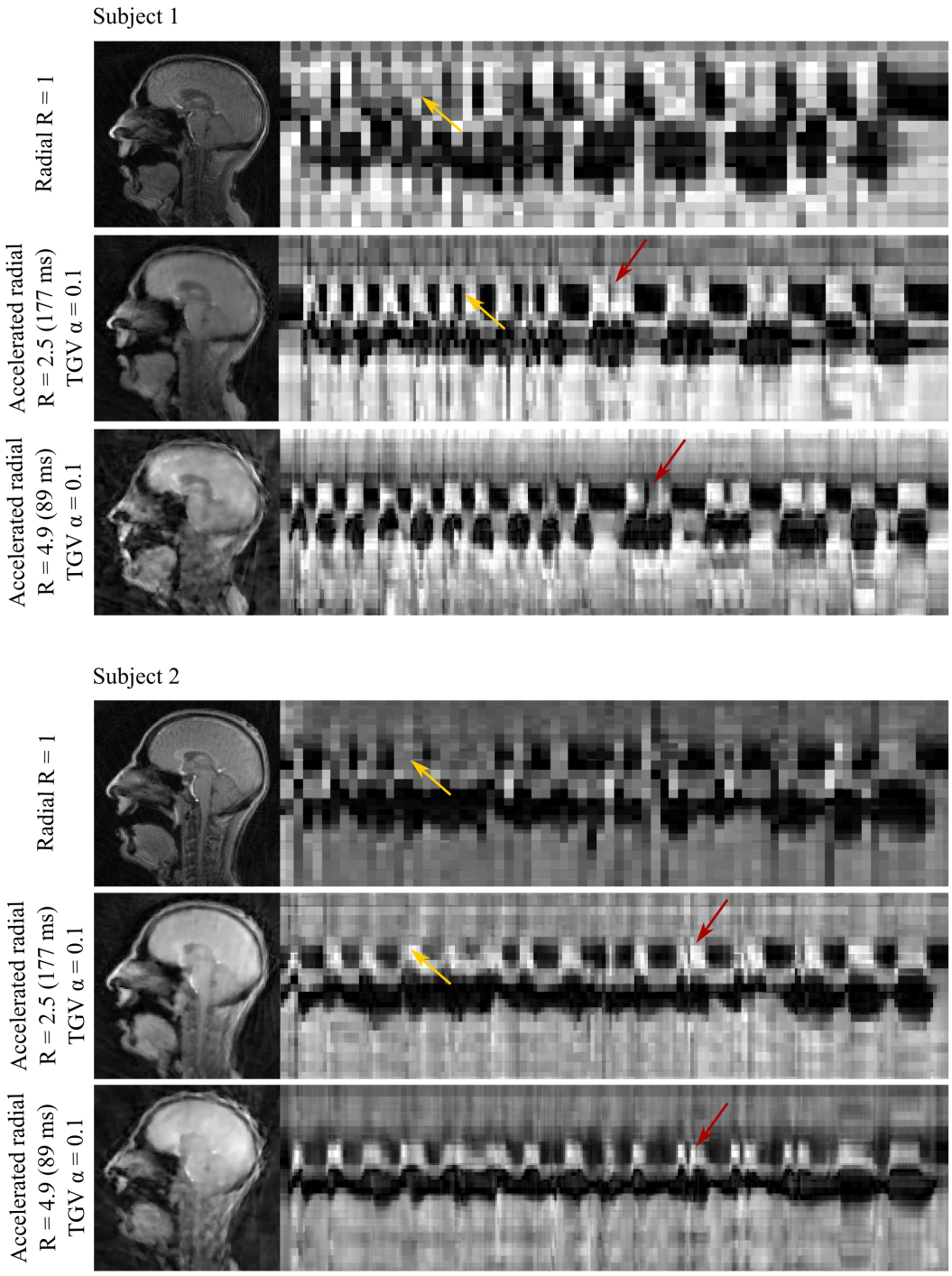


Figure 5.4: Real-time MRI of speech for two subjects of the cohort with fully sampled (R = 1, 108 radial projections) and accelerated (R = 2.5 and R = 4.9) reconstructed with iterative TGV regularised at $\alpha = 0.1$.

	CNR velum/air	**Temporal Sharpness**	**Spatial Sharpness**
$\alpha = 0$	9.82 (3.36)	0.11 (0.03)	0.07 (0.03)
$\alpha = 0.1$	14.42 (2.13)	0.17 (0.09)	0.07 (0.03)
p-value	< 0.05	0.17^{NS}	0.96^{NS}

Table 5.2: Mean and standard deviation measurements of CNR velum/air cavity, velum/air cavity sharpness in the temporal direction and velum/air cavity sharpness in the spatial direction across the entire cohort. Measurements were done on accelerated data sets (R = 2.5) reconstructed with L2 norm reconstruction (no TGV regularisation applied) and with the selected TGV regularisation of $\alpha = 0.1$. p-value refers to one-way ANOVA analysis. NS-not significant.

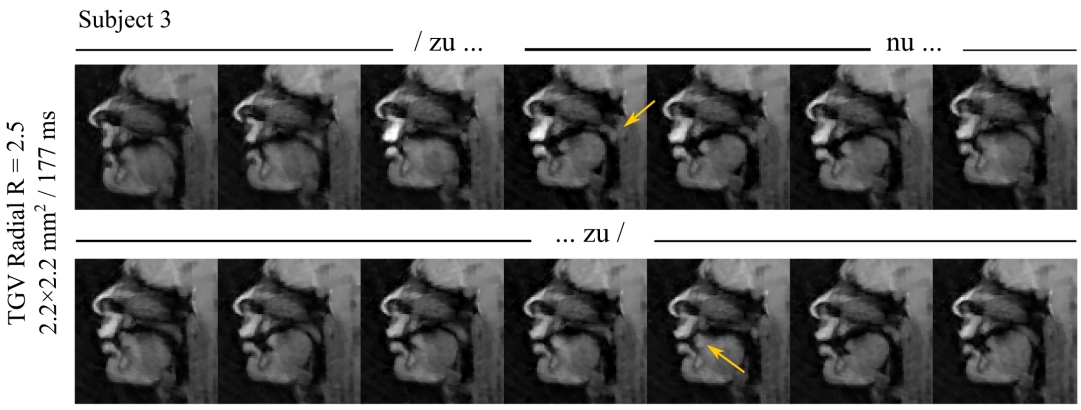


Figure 5.5: Rt-MRI data set at accelerated radial R = 2.5 ($\sim$ 7 fps) reconstructed with TGV during the phonation of the sound /zu-nu-zu/. Consecutive dynamic frames from the data set are shown where the elevation of the velum is visible for closure of the velopharyngeal port during /zu/ (yellow arrow) followed by a lowering of the velum for the vocalisation of /nu/ and again closure for the last /zu/.

5.3.3 Implementing GA sampling with view sharing TGV

An example of real-time speech MRI acquired with highly accelerated radial uniform and golden-angle acquisition, reconstructed with TGV regularisation, are shown in Figure 5.6. Data sets from two subjects of the cohort are shown at equivalent spatial and temporal resolutions (2.2×2.2 mm^2 and $\sim$11/12 fps), obtained with either high acceleration of conventional uniform radial sampling or view-sharing step of a golden-angle sampling. High acceleration of conventional radial data sets was achieved by acquiring 22 radial projections (R = 4.9), thus translating into a native frame rate of 89 ms per dynamic frame.

An increased display frame rate could also be achieved with the use of a golden-angle radial acquisition and selection of a view-sharing step equivalent to the highly under-sampled data (22 radial projections). A visually-selected TGV regularisation ($\alpha = 0.2$) was applied to this reconstruction after view-sharing between dynamic frames. Intensity-time profiles obtained from GA radial data sets, acquired with 43 radial projections and reconstructed with a view-sharing step of 22 radial projections, presented improved visual definition of closure events when compared to the speech task performance of data acquired with native 22 radial projections (Figure 5.6 yellow arrows).

5.4 Discussion

The feasibility of an iterative TGV-regularised SENSE-based reconstruction was assessed for the visualisation and assessment of velopharyngeal motion. Real-time MRI data of natural speech was successfully acquired in all four subjects, reconstructed with on-the-fly scanner vendor available gridding and additionally exported for offline TGV reconstruction.

In chapter 4, a comparison and evaluation of commercially available Cartesian and non-Cartesian rt-MRI methods was presented. Use of such widely available methods can be a starting point to image velopharyngeal motion and potential

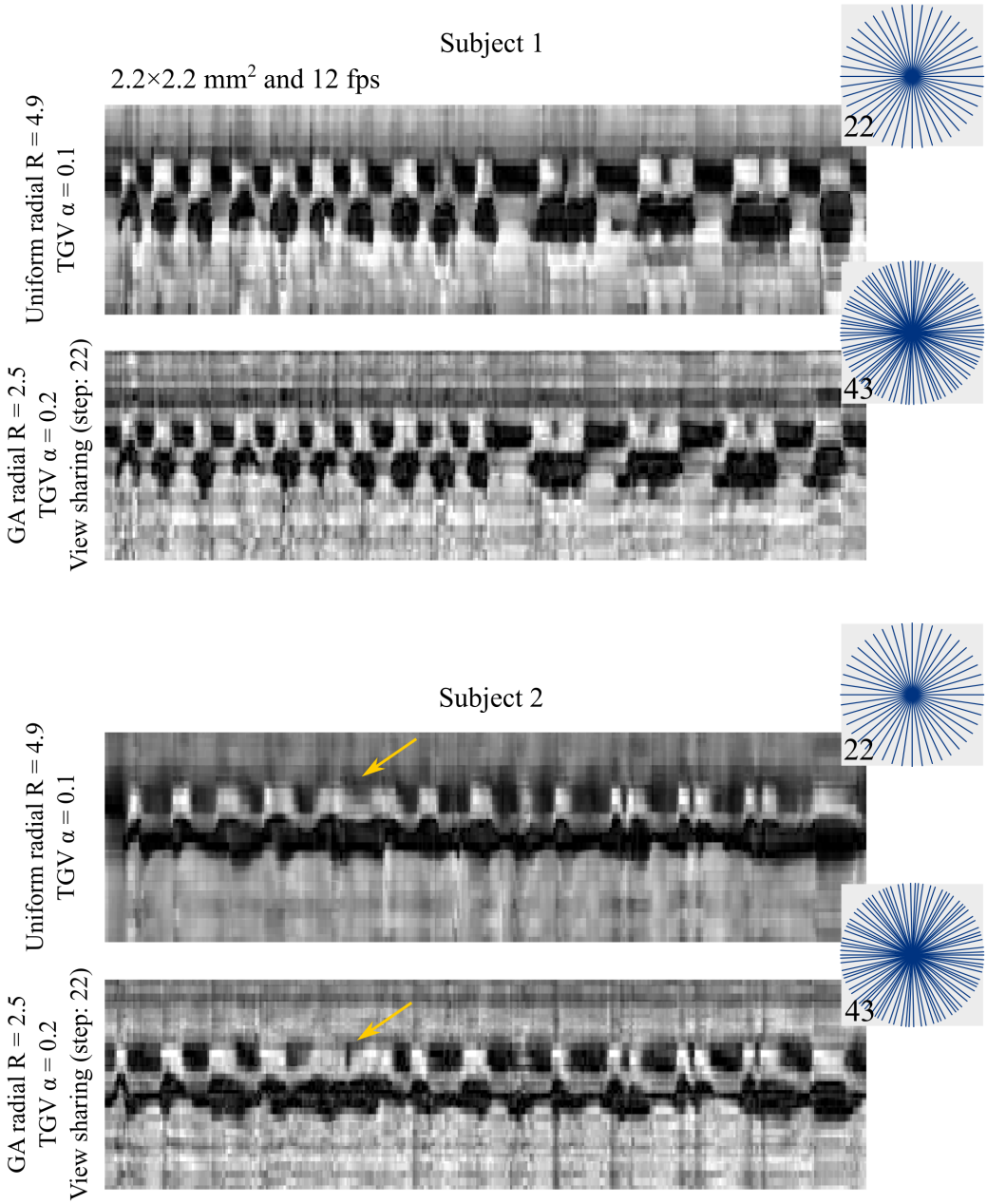


Figure 5.6: Rt-MRI of speech acquired with (a,c) conventional radial sampling (22 radial projections, R = 4.9) reconstructed with TGV $\alpha = 0.1$ and (b,d) with golden-angle radial scheme (43 radial projections, R = 2.5) reconstructed with TGV and view-sharing step of 22 radial projections. Improved depiction of temporally closer together closure events was noticeable with the GA view-sharing TGV scheme (yellow arrows).

future clinical translation. However, the true potential of non-Cartesian imaging can be further explored with less standard reconstruction methodologies.

One possible solution to this is to under-sample k-space and reduce total time of acquisition, thus increasing temporal resolution. Non-Cartesian under-sampling of k-space leads to less detrimental aliasing artefacts than coherent aliasing in under-sampled Cartesian k-space (refer to section 2.5). Thus, lower degrees of non-Cartesian k-space under-sampling may still be tolerated, as aliasing artefacts may not hamper image quality beyond clinical meaning. However, in cases of higher acceleration degrees, novel methods of non-Cartesian reconstruction can be considered to recover image quality and reduce detrimental aliasing artefacts. In this experimental chapter, I have investigated the use of iterative CG-SENSE based reconstruction regularised with a TGV constraint, a first application of this method to dynamic velopharyngeal imaging.

Firstly, and as in any SENSE-based method, estimation of the coil sensitivities is necessary. In this experimental work (similarly to the original TGV implementation), coil maps were iteratively estimated from accelerated raw radial data sets. One advantage of this is that no additional scans (e.g. external estimation of coil maps with a Q-body coil) are necessary in order to reconstruct raw data with this method. However, some aliasing artefacts are present in the estimated coil maps (Figure 5.1), which can later hamper final image quality, particularly at higher acceleration degrees.

Selection of the regularisation weight (α) was performed empirically by visual inspection of the dynamic rt-MRI data sets and additional qualitative measurements (CNR and sharpness at the velum/air boundary). As expected, when no regularisation was considered ($\alpha = 0$), a standard L_2 norm iterative reconstruction is performed and final reconstructed images presented increased random noise (and thus decreased CNR performance) and increased presence of aliasing artefacts. On the other hand, over-weighting of the TGV regularisation term (extreme case of $\alpha = 1$) leads to the removal of essential image features (vocal tract shape) and

over-smoothing of the intensity-time profiles. Therefore, α was selected based on visual assessment such that it minimised aliasing artefacts while still maintaining good segmentation of the vocal tract shape. For the present data sets at 2.2×2.2 mm^2, an optimal $\alpha = 0.1$ was chosen. Improved CNR at the velum/air boundary was observed for data sets reconstructed with optimal regularisation $\alpha = 0.1$ compared to no TGV and over TGV regularisation, $\alpha = 0$ and $\alpha = 1$ (21.15 ± 2.31 vs. 10.01 ± 2.16 and 14.98 ± 1.20).

Accelerated radial data sets at R = 2.5 ($\sim$7 fps) and R = 4.9 ($\sim$11 fps) reconstructed with $\alpha = 0.1$ demonstrated adequate sampling of the velopharyngeal motion. For data sets acquired at $\sim$ 7fps, temporal resolution was just about adequate for the assessment of velopharyngeal motion during natural speech [41]. Additionally, further improvement in sampling the velopharyngeal closure pattern was observed with highly accelerated data at $\sim$ 11 fps. However, an increase in temporal resolution resulted in diminished final image quality in the dynamic data sets (increased presence of aliasing artefacts), although an overall assessment of the vocal tract shape is still possible. Feasibility of the selected TGV regularisation was also assessed by measurements of CNR and sharpness at the velum and air cavity boundaries for the entire cohort. The optimised TGV regularisation $\alpha = 0.1$ provided an increase of about 47% in CNR performance compared to non-regularised iterative reconstruction.

As previously shown, highly accelerated data sets provided improved temporal resolution for the assessment of velopharyngeal closures (R = 4.9 at $\sim$ 11 fps), however hampered by decreased image quality. An alternative method to increase final display temporal resolution was to consider a moderately accelerated native frame rate ($\sim$ 7fps, minimum for velum imaging) and increase final frame rate by means of a view-sharing scheme. As explored in the beginning of this chapter (section 5.1.3), the use of a golden-angle radial trajectory allows to uniformly sample k-space, independently of the number of selected radial projections. Therefore, when a view-sharing scheme is used (that is, when data is to be "shared" between

acquired images), a uniform sampling of k-space is always present, thus reducing the aliasing artefacts level in the final image.

Rt-MRI of two subjects of the cohort were compared at similar spatial and temporal resolutions obtained with either highly accelerated conventional radial trajectory (22 projections) and GA sampling with view-sharing (43 projections reconstructed with a sharing step of 22 projections). In this case, the implementation of GA sampling added flexibility to the selection of final display temporal resolution (i.e. any other sharing step could have been chosen). Therefore, although data sets were acquired at $\sim$ 7fps (43 radial projections), information could be displayed to a much higher temporal resolution, equivalent to highly accelerated data sets (22 projections, $\sim$ 12 fps), while yielding adequate image quality and diminished temporal blurring and/or distortion due to aliasing introduced by the high acceleration factors, as seen in the conventional R = 4.9 radial acquisition (Figure 5.6).

One main advantage of this reconstruction scheme is that no additional data is required (e.g. fully-sampled calibration data or external coil maps estimation) in order to reconstruct rt-MRI data sets. On the other hand, required times for offline processing and reconstruction of raw data constitute the main limitation of this work. With this particular set-up, reconstruction of one dynamic frame took approximately 60-80 seconds depending on degree of radial acceleration and number of CG loop iterations. This meant that for the selected speech task sample with accelerated radial scans at R = 2.5 and R = 4.9, reconstruction times ran between 3 and 6 hours. Reconstruction and processing times could be further improved with the use of GPU processing and a CUDA implementation [74]. However, original descriptions of CUDA implementation reported reconstruction times in the order of 60 to 100 seconds per image, which in the case of hundreds of dynamic frames (such as speech rt-MRI data sets) would still be prohibitive for future clinical translation of speech assessment.

Therefore, the work undertaken in the following and last experimental chapter

of this thesis will focus on the use of a different reconstruction methodology (a GRAPPA based method) in order to overcome this intrinsic limitation of iterative based reconstruction methods.

5.5 Summary

In this chapter, feasibility of an accelerated conventional and GA radial acquisition reconstructed with TGV-regularised method was established for the visualisation and assessment of velopharyngeal motion. It was established that TGV-regularised reconstruction with an accelerated radial trajectory could be used to adequately sample velopharyngeal motion. However, some constraints, mainly the required times for offline reconstruction and processing, deemed this methodology prohibitive for future clinical translation in the assessment of VPI and speech disorders.

Results presented in this experimental Chapter have been partially published in:

A C Freitas, M Ruthven, R Boubertakh, M E Miquel. Real-time speech MRI: commercial Cartesian and non-Cartesian sequences at 3T and feasibility of offline TGV reconstruction to visualise velopharyngeal motion. *European Journal of Medical Physics*, 2018. (*in press*)

Chapter 6

Improved 2D rt-MRI of speech using accelerated radial tt-GRAPPA in velopharyngeal imaging

This chapter demonstrates the feasibility of accelerated radial acquisition and parallel imaging to improve real-time MRI of speech. It explores the use of a GRAPPA based method with through-time calibration to reconstruct under-sampled non-Cartesian data, and its application to dynamic imaging of natural speech.

6.1 Introduction

6.1.1 Cartesian GRAPPA: theory overview

A common parallel imaging method used to accelerate image acquisition is GRAPPA [30]. In addition, non-Cartesian GRAPPA can also be derived from the traditional principles of Cartesian GRAPPA [26].

As previously covered, skipping phase-encoding lines in Cartesian k-space re-

duces total scan time. However, skipping k-space lines leads to aliasing in image space, thus hampering image quality. Unlike image-based methods (such as SENSE) that attempt to reconstruct coil-combined images and require explicit knowledge of coil sensitivity information, GRAPPA estimates the missing data points in k-space on a coil-by-coil basis. In this case, each missing data point, commonly known as *target point*, is estimated from the combination of a kernel of data points and a set of GRAPPA weights. The chosen kernel defines the distribution between contributing known k-space data points (*source points*) and the unknown data points (*target points*). Although many kernel sizes and geometries can be used, the following description in this chapter relies on a 2×3 kernel, as exemplified in Figure 6.1. Then, all source points are multiplied by the estimated GRAPPA weights and summed in order to calculate the resulting target point. Considering a vector describing all target points in a coil i ($S_{target,i}$), all source points ($S_{source,i}$) and GRAPPA weights (ω), this can be described in matrix form as [26]:

$$S_{target,i} = \omega . S_{source,i} \tag{6.1}$$

However, ω needs to be estimated first. In the Cartesian case, and due to the uniformity of k-space sampling, GRAPPA is said to be auto-calibrated. This means that the weights (ω) can be estimated from a fully-sampled number of k-space lines (known as ACS: auto-calibration signal) where sufficient combinations of both source and target data points are known (Figure 6.1). Thus, by means of a pseudo-inverse operation, an accurate estimation of ω can be calculated:

$$\omega = pinv(S_{source,i}) . S_{target,i} \tag{6.2}$$

The calibration equation is fully determined when:

$$n_{rep} \geq n_c \times n_{points} \tag{6.3}$$

That is, when the number of repetitions of the kernel over k-space (n_{rep}) is greater than the number of coil elements times the number of source data points in the kernel ($n_c \times n_{points}$). If n_{rep} is insufficient, weights estimation is mostly influenced by noise and the final reconstruction of the image inaccurate.

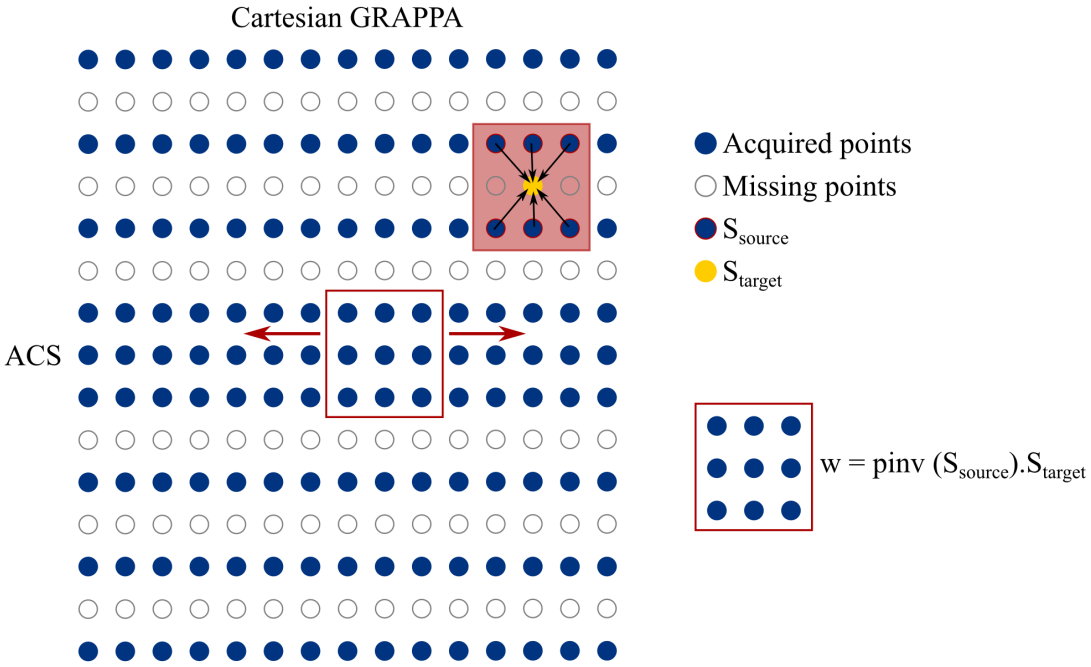


Figure 6.1: Schematic representation on conventional Cartesian GRAPPA (adapted from [26]). In the Cartesian case, GRAPPA is said to be auto-calibrated as GRAPPA weights can be estimated from a set of centre fully-sampled k-space lines known as the ACS region. Thus, the weights may be estimated by a pseudo-inverse operation of all known and unknown k-space data points in the kernel.

6.1.2 GRAPPA in non-Cartesian k-space

The main difference between conventional Cartesian GRAPPA and non-Cartesian GRAPPA is the non-uniformity of k-space coverage. Because non-Cartesian trajectories sample k-space in a non-uniform way, GRAPPA kernels in non-Cartesian k-space are uniquely shaped and the relation between source and target data points

is distinct for each kernel. In theory, each target data point in a non-Cartesian sampling (e.g. radial) presents a unique combination of target and source data points, which would mean that a GRAPPA weight had to be calculated for each target point [26]. In addition, each unique kernel geometry only occurs once per k-space and therefore there is no accurate way to calculate the GRAPPA weights.

An initial solution to non-Cartesian GRAPPA was presented by acquiring a fully-sampled data set (ACS) and defining k-space segmentation. In this case, k-space must be divided into multiple large segments; large enough that the kernel can occur multiple times inside the segment and thus, accurately calculate the weights for that portion of k-space. However, choice of segment size can still present many challenges. A larger segment might allow for many kernel occurrences but calculated weights are less accurate. Using smaller segments might ensure that the kernels present similar geometry and a more accurate calculation of the GRAPPA weights, however, too small a segment might not encompass the necessary minimum occurrence of the kernels for the weight calculation equations to be over-determined.

6.1.3 Through-time GRAPPA

In order to overcome previously explored limitations of non-Cartesian GRAPPA with k-space segmentation, a novel calibration scheme was proposed by Seiberlich et al., through-time GRAPPA or tt-GRAPPA [12]. This method has been mainly used in dynamic MR imaging of the heart where the consecutive acquisition of the same set of image data (dynamic frames) is naturally performed. In this case, rather than relying on the multiple repetitions of the kernel through-k-space (i.e. inside the larger selected segment), weights can be calculated from the multiple repetitions of the kernel through time (i.e. along the different dynamic frames). Since each target point presents a unique kernel geometry that is repeated through time several times, i.e. as many times as dynamic frames are acquired, GRAPPA weights can be more accurately calculated (refer to Figure 6.2). One pitfall to

this technique is that it requires the acquisition of multiple fully-sampled dynamic frames pre- or post-scanning which can be time consuming. However, due to the more accurate estimation of the GRAPPA weights, higher acceleration factors can usually be achieved with optimal minimisation of aliasing artefacts [12].

As previously stated, one disadvantage of this technique to direct translation to clinical applications is the lengthy acquisition of calibration data. Seiberlich et al. described that a fully-sampled radial ACS could take up to 2.82 minutes of continuous acquisition (400 dynamic frames × 144 radial spokes per frame × TR 2.92 ms). One solution to reduce total acquisition time and preserve reconstruction quality is to employ a hybrid GRAPPA scheme. Hybrid GRAPPA allows to reduce the number of necessary calibration frames by performing calibration both through-k-space and through-time.

Hybrid and purely tt-GRAPPA have been initially described in real-time, non-gated, free-breathing cardiac imaging with both radial and spiral acquisitions [12, 58]. Initial implementation with a 6-fold acceleration radial trajectory allowed for real-time cardiac imaging at a temporal resolution of 36 ms and a spatial resolution of 1.56×1.56 mm^2. The same reconstruction scheme was later used to reconstruct highly accelerated spiral trajectories. As previously noted, spirals are inherently efficient trajectories, allowing to cover k-space in approximately 60% of the time of an equivalent radial trajectory. Thus, a much lower acceleration factor (R = 6) was possible in spiral trajectories compared to radial trajectories (R = 9) to achieve the same temporal resolution, about 35 ms [58]. A high acceleration of spiral trajectories at R = 12 could also be accurately reconstructed with no major presence of aliasing artefacts at 18 ms per dynamic image.

Through-time non-Cartesian GRAPPA was later translated to other clinical applications, such as 3D MR renal angiography [79] and T_1 mapping of the abdomen [80].

In addition, Lingala et al. has recently, during the experimental work undertaken in this thesis, applied non-Cartesian tt-GRAPPA to real-time MRI of

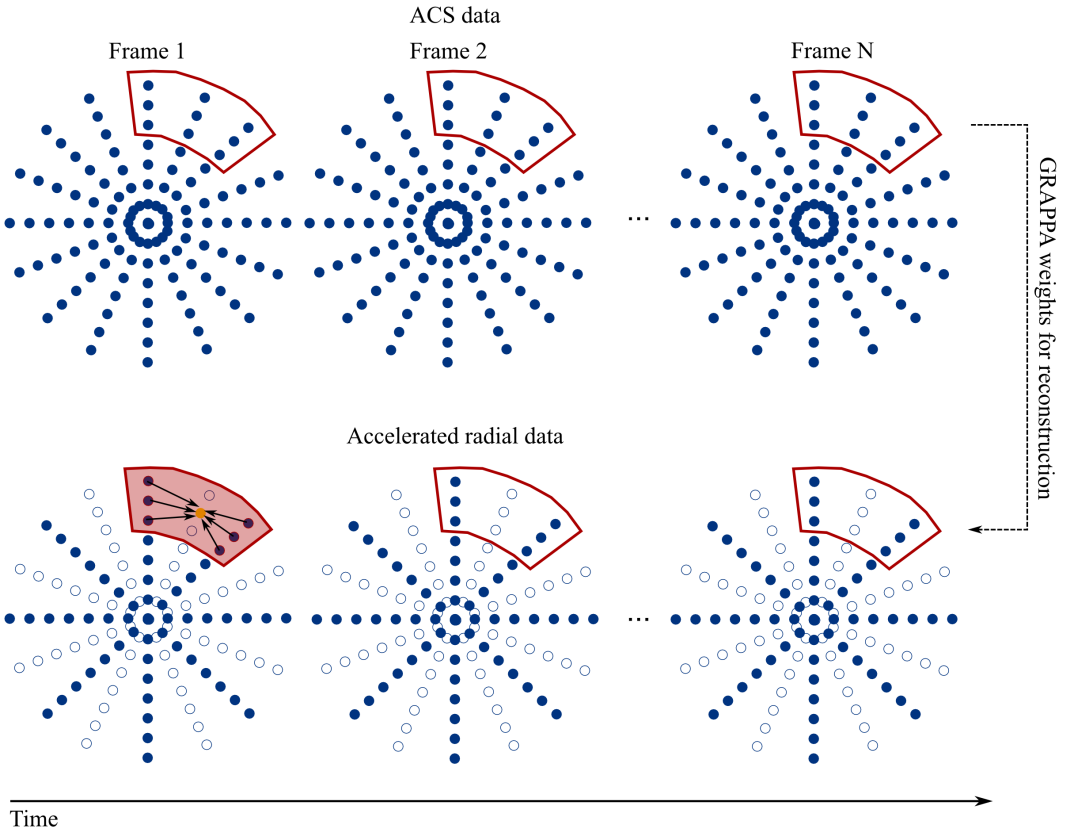


Figure 6.2: Schematic representation of through-time GRAPPA reconstruction of accelerated radial data sets. A kernel of source and target points is repeated through-time in the ACS data, as many times as there are dynamic frames in the data set (N frames). After the GRAPPA weights are calculated, missing target points in the accelerated data sets can be estimated. Adapted from [12].

speech [59, 81]. Spiral GRAPPA data was assessed for a range of speech tasks from natural speech to rapid articulatory motion of beat-boxing. Spiral GRAPPA was observed to maintain good temporal depiction of the articulators motion at acceleration folds up to R = 4. Additionally, through-time spiral GRAPPA improved temporal depiction of rapidly moving vocal tract articulators when compared to the blurred imaging of a view-sharing reconstruction. An optimised reconstruction protocol was set up with a R = 4 spiral acceleration factor (18 ms per dynamic frame) and a 1×1×150 (azimuthal×readout×time directions) GRAPPA calibra-

tion scheme. This means that the required kernel was repeated 1×1 in-plane through-k-space (i.e. with no in-plane segmentation) and ×150 through-time (dynamic frames). Reconstruction times with spiral GRAPPA were reported 8 times slower than view-sharing reconstruction. Furthermore, many improvements have been suggested by the author as not yet explored, such as the application of radial trajectories (better suited at higher field strengths as spirals are particularly prone to off-resonance effects) and the use of hybrid calibration.

In conclusion, non-Cartesian tt-GRAPPA is a promising tool for the reconstruction of dynamic imaging. However, its feasibility in other clinical applications is still unexplored and there is no current application of the technique to assess velopharyngeal motion in speech. In this chapter, an extensive evaluation of through-time/through-k-space GRAPPA to speech imaging is performed. I set out to address some of the open questions raised by previous authors, such as its application to radial trajectories and hybrid calibration, as well as a first application of this methodology to the assessment of velopharyngeal motion and closure pattern.

6.2 Methodology

6.2.1 MR image acquisition and tt-GRAPPA implementation

Experiments were performed on a Philips Achieva 3 T Tx clinical scanner (Philips Healthcare, Best, the Netherlands) R 3.3 with a 16-channel neurovascular coil.

A spoiled-GRE radial trajectory was implemented with imaging parameters as follows: FOV = 190×190 mm^2, base acquisition matrix = 100×100, TE/TR = 2.06 ms / 4.53 ms, flip angle (α) = 10 degrees, 2D single-slice 10 mm thick, in-plane spatial resolution = 1.9×1.9 mm^2. Temporal acceleration of radial data was accomplished by under-sampling of k-space, i.e. by reducing the number of radial

projections acquired per base matrix. As previously explained in Chapter 2, the minimum number of radial spokes required to obtain an aliasing-free image is given by the Nyquist criterion as $\frac{\pi}{2} \times N_{matrix}$. Thus, for these experiments, 157 radial spokes would in theory be necessary to obtain an aliasing-free image. However, the implementation of tt-GRAPPA requires the number of sub-sampled radial spokes to be an integer divisor of the number of available radial spokes in the calibration data (fully-sampled). As a low degree of sub-sampling can usually be tolerated, calibration data (ACS) at 100 spokes was considered instead to allow for an easier under-sampling of radial spokes (that is because 40 or 20 radial spokes divide 100 in an integer number but not 157). Therefore, in this experimental work, the degree of acceleration (R) was defined in comparison to the fully-sampled Cartesian case, i.e. R = 1 corresponds to 100 radial spokes acquired in a 100×100 base matrix. A previous implementation of non-Cartesian tt-GRAPPA has prescribed ACS data to be acquired with up to 400 dynamic frames [12]. However, translation of this technique to speech imaging is hampered by the duration that a standard healthy adult can sustain continuous natural speech without the need to move. In order to investigate the effect of the number of dynamic frames in the ACS data (corresponding to number of kernel occurrences through time) on final image quality, preliminary phantom work was also undertaken.

Following GRAPPA weights estimation, individual coil images were generated with an open-source NUFFT toolbox [24] and later combined with a sum-of-squares method. Offline reconstruction of image data was implemented within MATLAB (The Mathworks, Inc., Natick, MA, R2014b) on a 2.5 GHz Intel Core i7, 16 Gb computer.

6.2.2 Phantom and *in-vivo* experiments

Preliminary phantom work was firstly undertaken, acquiring ACS data (R = 1) with 400 dynamic frames. Data sets were retrospectively sub-sampled by a 5 and 10-fold degree (R = 5 and R = 10) and reconstructed with tt-GRAPPA to invest-

igate the effect of the number of calibration frames on final reconstructed image quality. Sub-sampled data was then reconstructed using a fixed in-plane segment size (in this case, readout×projection: 1×1) and a varying number of through-time calibration (i.e. varying number of dynamic frames used for reconstruction). As previously explored, the minimum number of kernel repetitions required for an accurate reconstruction is given by the number of coils times the number of data points in the kernel. For all the experimental work here described, the minimum kernel occurrences is then given by $16\times(2\times3) = 96$. Thus, 100 temporal frames was set as the minimum value used and 400 as the maximum value. Using a step of 25, the number of temporal frames used for GRAPPA weight estimations was varied as follows: [100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400].

Four healthy adult volunteers (four males, mean age 37 years) were recruited for this experimental work. Subjects were imaged in the supine position while performing a speech sample consisting of counting from 1 to 10, non-sense nasal phonation (/za-na-za/, /zu-nu-zu/ and /zi-ni-zi) and sustained phonation of vowels (/a/ and /i/) at a natural rate. Image data was acquired at multiple acceleration factors such as R = 2, 4, 5 and 10. In addition, calibration data was also acquired pre-scanning with 200 dynamic frames (results derived from preliminary phantom work). Further details on acquisition parameters are described in Table 6.1. During the acquisition of calibration data, the subjects were asked to repeat the speech task until acquisition was finished. Total acquisition time for ACS data was approximately 1 minute and 30 seconds (4.53 ms × 100 radial projections × 200 frames).

Due to the lengthy acquisition time required to acquire ACS data (up to ∼1.50 minutes) and consequent limitations to future clinical translation, investigation of a hybrid GRAPPA calibration was of great interest in this experimental work. Thus, a solely through-time calibration (1×1×Number of temporal frames) and a hybrid through-time/through-k-space calibration were investigated. A through-k-space segmentation to allow for sufficient kernel repetitions was proposed with

Temporal resolution	Spatial resolution (mm^2)	Matrix	N_{spokes}	R
454 ms (∼2.3 fps)	1.9×1.9	100×100	100	1
227 ms (∼4.4 fps)	1.9×1.9	100×100	50	2
110 ms (∼9.0 fps)	1.9×1.9	100×100	25	4
91 ms (∼11.0 fps)	1.9×1.9	100×100	20	5
45 ms (∼22.1 fps)	1.9×1.9	100×100	10	10

Table 6.1: Acquisition parameters reported at different acceleration folds (R) of radial acquisition for tt-GRAPPA experiments.

in-k-space repetitions as (readout × projection: 2×1, 4×2 and 8×1) and temporal segmentation was proposed using temporal frames ranging from 5 to 200 ACS temporal frames. Dimensions of through-k-space segmentation were chosen following initial reports of tt-GRAPPA [12], as similar kernel configurations within the segment seem to hold better when the segment is larger in the readout direction. Visual assessment and measurements of error were undertaken to assess an optimal calibration scheme for speech imaging.

6.2.3 Data analysis

Accelerated data sets were evaluated against a set of reference data with measurements of root mean squared error (RMSE). RMSE was defined as follows:

$$RMSE = \sqrt{\frac{\sum(Img - Img_{ref})^2}{n}} \tag{6.4}$$

Where Img is the accelerated reconstructed final images, Img_{ref} is the image data used as reference and n the total number of images. Since no true reference data exists in this case, image data reconstructed with the largest number of calibration frames was used as reference data. RMSE was calculated across all dynamic frames of each data set.

Measurements of CNR of the velum and neighbouring air cavity (previously described in Chapter 4) in a short section of the intensity-time displays of velopharyngeal motion were also undertaken. ROIs were selected in the velum and neighbouring air cavity in the intensity-time displays during the phonation of the sound /zi-ni-zi/. Measurements of temporal and spatial sharpness at the velum and air cavity boundaries were also undertaken as previously described in Chapter 5.

Statistical analysis was undertaken using SPSS software (v. 22, IBM, New York). For continuous variables of multiple measurements, such as comparing CNR performance between reconstructed images, a repeated-measures one-way analysis of variance (ANOVA) was undertaken with significance level of 0.05. Post-hoc paired t-test with Bonferroni correction for multiple testing was used to identify significant pairs.

6.3 Results

6.3.1 Preliminary phantom work

Measurements of RMSE were undertaken in order to compare radial under-sampled images reconstructed with tt-GRAPPA and the original fully-sampled calibration images. RMSE measurements against varying the number of temporal calibration frames, for a 5 and 10-fold acceleration degree, are shown in Figure 6.3. Results showed a decrease in measured RMSE between image data sets reconstructed with 100 and 125 (1×1) temporal calibration dynamic frames. Further analysis of the RMSE curve from 125 to 400 temporal calibration frames seems to indicate that negligible changes in reconstruction quality occur beyond using 200/225 temporal calibration frames. Thus, a value of 200 temporal calibration frames was selected for all following *in-vivo* experiments.

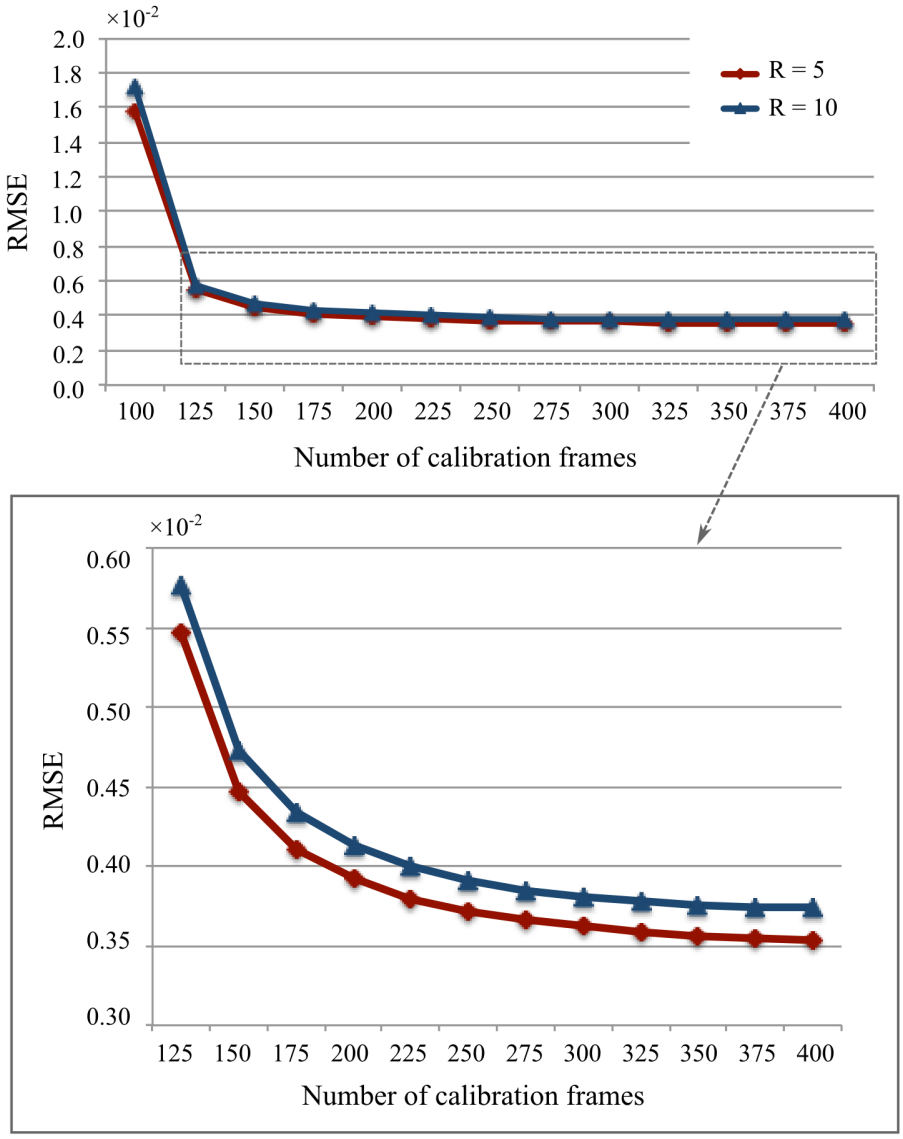


Figure 6.3: Measurements of RMSE to investigate the effect of tt-GRAPPA reconstruction using different numbers of calibration dynamic frames. Data was retrospectively sub-sampled from the calibration data fully sampled (R = 1) to a 5-fold (red) and 10fold (blue). Sub-sampled data sets were reconstructed with kernel repetitions at 1×1×N, N being the number of calibration temporal frames, varying from 100 to 400. RMSE was then measured comparing the reconstructed sub-sampled data sets at R = 5 and R = 10 to the fully sampled original data sets.

6.3.2 *In-vivo* validation for speech imaging

In-vivo validation of the use of tt-GRAPPA for accelerated speech imaging was undertaken. In order to evaluate image error due solely to reconstruction accuracy of tt-GRAPPA, data sets were retrospectively sub-sampled and simulated at multiple folds of acceleration (R = 2, 4, 5, 10 and 20) from the fully-sampled calibration data (R = 1, 100 projections). Data sets were then reconstructed using a through-time GRAPPA calibration of 1×1×200. RMSE measurements and difference images demonstrating overall artefact level are shown in Figure 6.4. As seen in Figure 6.4 b, presence of artefacts solely due to tt-GRAPPA were mainly concentrated in the region of the lips and vocal tract resulting in a higher RMSE in the selected ROI over the velopharyngeal region (Figure 6.4 (a) blue) than measured over the overall image (red). However, even at a higher radial acceleration of R = 5, the level of image distortion/error remained equivalent to low-acceleration images (R = 2).

Figure 6.5 shows real-time speech MRI of two subjects repeating the selected speech task, acquired with accelerated radial R = 5 (91 ms per frame) and reconstructed with zero-filled NUFFT (Figure 6.5 a,c) and tt-GRAPPA (Figure 6.5 b,d). Zero-filled NUFFT reconstruction showed increased image distortion due to aliasing (i.e. insufficient sampling of k-space data). In the case of radial trajectories, aliasing translates into sharp spoke-like artefacts in the final reconstructed image. On the other hand, reconstruction of the same data sets with tt-GRAPPA (through-time calibration: 1×1×200) provided imaging data with diminished aliasing artefacts present and clear definition of the vocal tract articulators. Additionally, it also demonstrated adequate temporal depiction of velopharyngeal motion as a clear identification of expected velopharyngeal closures corresponding to phonation sounds was possible in the intensity-time profiles (Figure 6.5 b).

Examples of tt-GRAPPA intensity-time profiles of the velopharyngeal motion from data sets acquired at several accelerated radial factors (R = 1, 2, 5 and

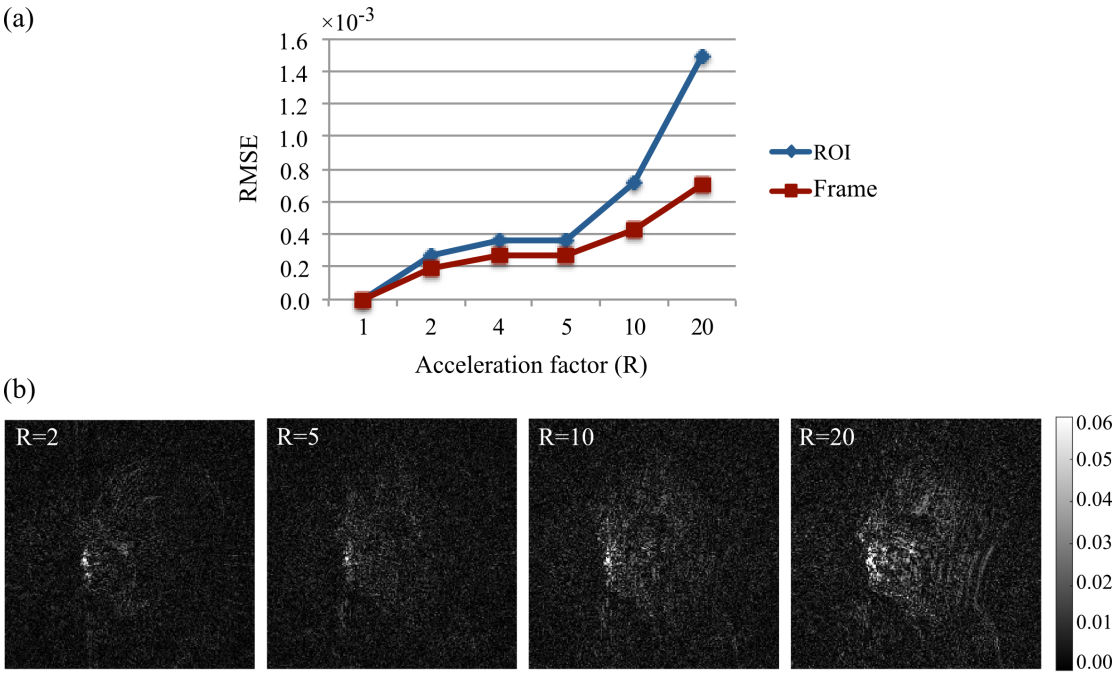


Figure 6.4: (a) Mean RMSE measurements from retrospectively sub-sampled radial data (R = 2, 4, 5, 10 and 20) compared to the fully sampled original data (R = 1). RMSE measurements were performed both in a ROI selected in the velopharyngeal region (blue) and the overall frame (red). (b) Difference images between sub-sampled and reference data corresponding to acceleration factors of R = 2, 5, 10 and 20.

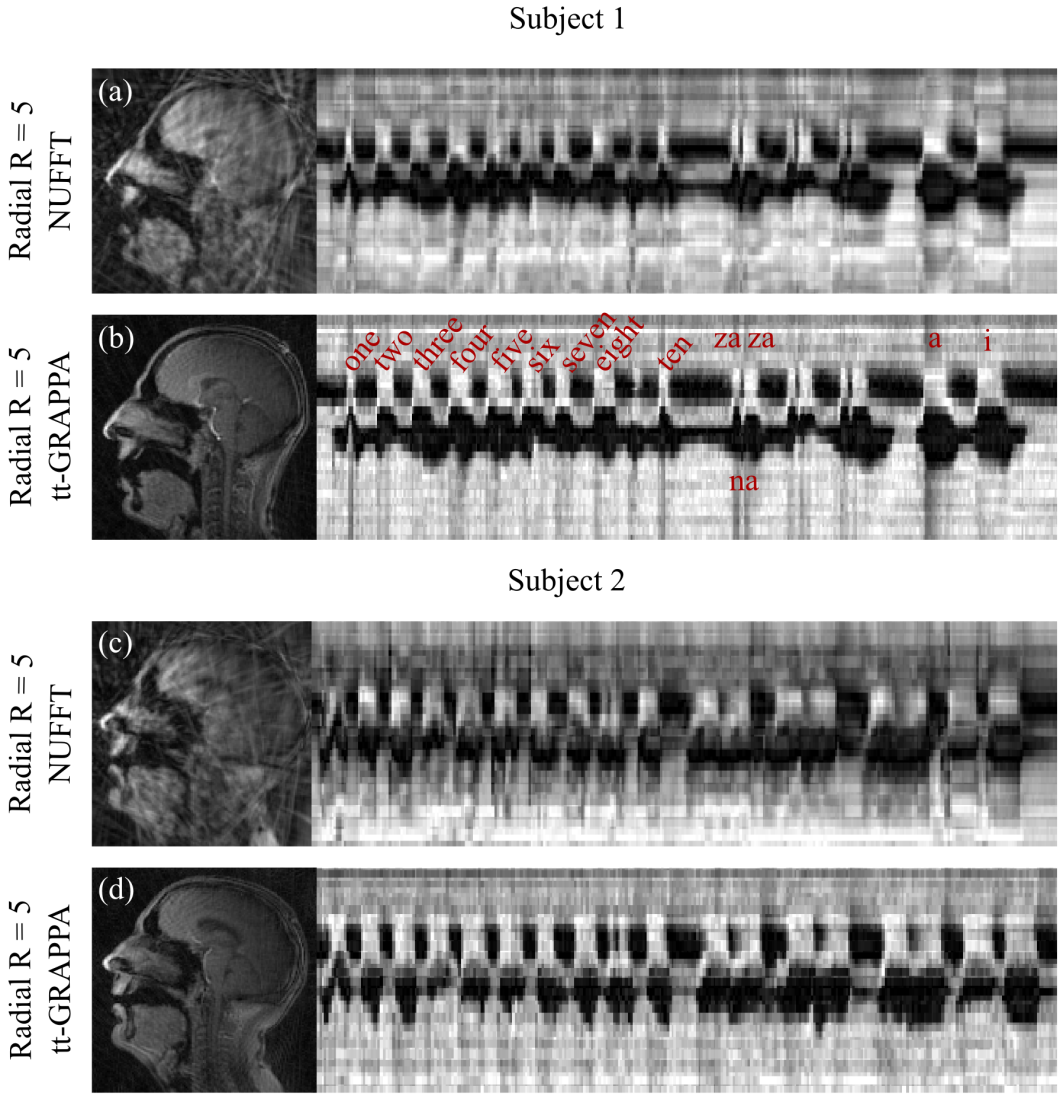


Figure 6.5: Real-time imaging of natural pace speech acquired at 5-fold radial acceleration (∼11 fps). Data reconstructed with (b,d) tt-GRAPPA at 1×1×200 and with (a,c) zero-filled NUFFT for comparison. Example of mid-sagittal dynamic frames are shown on the left side and intensity-time displays of velopharyngeal motion during phonation are shown on the right hand side. Position and motion of the velum according to the corresponding phonation sound can be identified in the intensity-time displays.

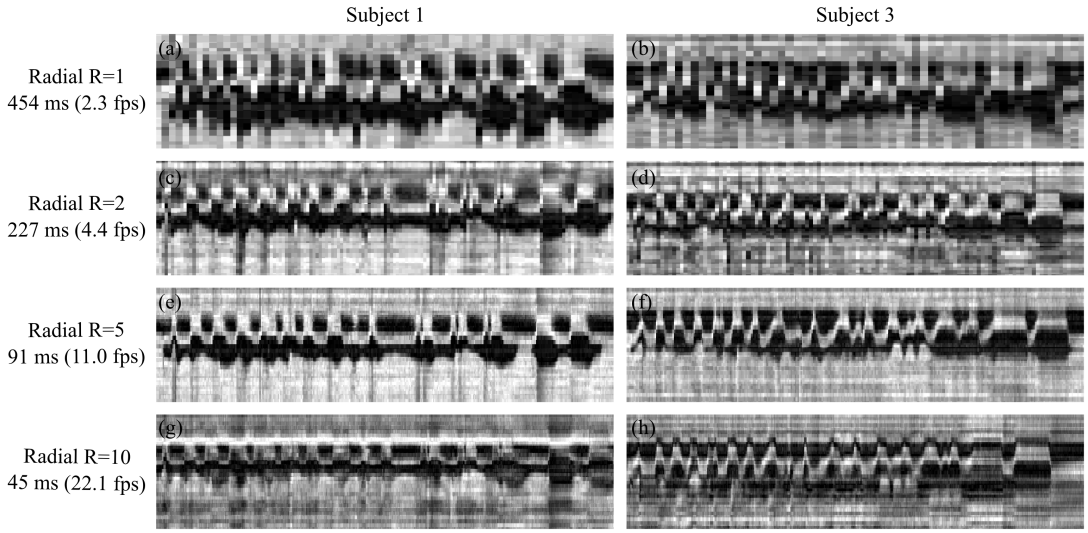


Figure 6.6: Through-time GRAPPA intensity-time displays of velopharyngeal motion from two subjects of the cohort derived from rt-MRI data sets acquired at radial acceleration factors of R = 1, 2, 5 and 10. Data sets correspond to temporal resolutions ranging from 454 ms up to 45 ms.

10) are shown in Figure 6.6. Results showed that at a low sampling rate, corresponding to the calibration data R = 1 (∼2.3 fps), velum motion representation was inadequate and closure events are blurred and not all identifiable. At a higher acceleration-fold (R = 2, 4.4 fps), the overall pattern of velopharyngeal motion can already be distinguished. However, velopharyngeal closures that might be more temporally closer together can still be blurred and non-identifiable (such as the cases of phonation of /za-na-za/). On the other hand, at a higher sampling rate R = 5 (∼11 fps), all expected velopharyngeal closure events are easily identified and with a crisp depiction. Lastly, temporal resolution could be further improved with a high acceleration fold of R = 10 (∼22.1 fps), although with final image quality hampered by decreased reconstruction quality.

6.3.3 GRAPPA calibration: through-time vs. hybrid

Through-time and hybrid through-k-space/through-time GRAPPA calibration schemes were investigated in the reconstruction of accelerated speech imaging. Figure 6.7 shows measurements of RMSE depending on in-plane k-space segmentation (1×1, 2×1, 4×2 and 8×1) and temporal kernel occurrences (5 to 200 temporal frames), as well as examples of mid-sagittal dynamic images.

Results showed that the higher the number of calibration frames used during reconstruction, the further the improvement in overall image quality. Visual assessment of accelerated radial R = 5 data sets reconstructed with 1×1×200 demonstrated optimal image quality for the assessment of velopharyngeal motion. Thus, because no true reference data exists in this case, images reconstructed with 1×1×200 were used as reference when calculating RMSE. If the number of dynamic frames is insufficient to accurately determine the GRAPPA weights, image quality decreases as reconstruction errors and higher dependency on noise are introduced. Visual assessment of image data sets reconstructed with insufficient kernel repetitions (e.g. 1×1×5) showed increased distortion from reconstruction errors (Figure 6.7 c). In cases where the weights equation is just weakly determined (such as 1×1×100 or 2×1×50), reconstructed images are borderline accurate but showed higher dependency on random noise. Because this distribution of overall high levels of noise increases RMSE compared to the reference data, this causes a peak in the RMSE plot curves (as seen in Figure 6.7 b). In addition, a hybrid through-time/through-k-space calibration scheme (e.g. 4×2×75) allowed maintaining adequate image quality (and equivalent RMSE) while reducing the required number of calibration frames (i.e. 75 instead of 200 dynamic frames).

Examples of mid-sagittal dynamic frames and corresponding intensity-time profiles of velopharyngeal motion reconstructed with different GRAPPA calibrations are shown in Figure 6.8. As previously explored, both the cases of 1×1×5 and 4×2×10 carry insufficient kernel repetitions for the GRAPPA weights to be

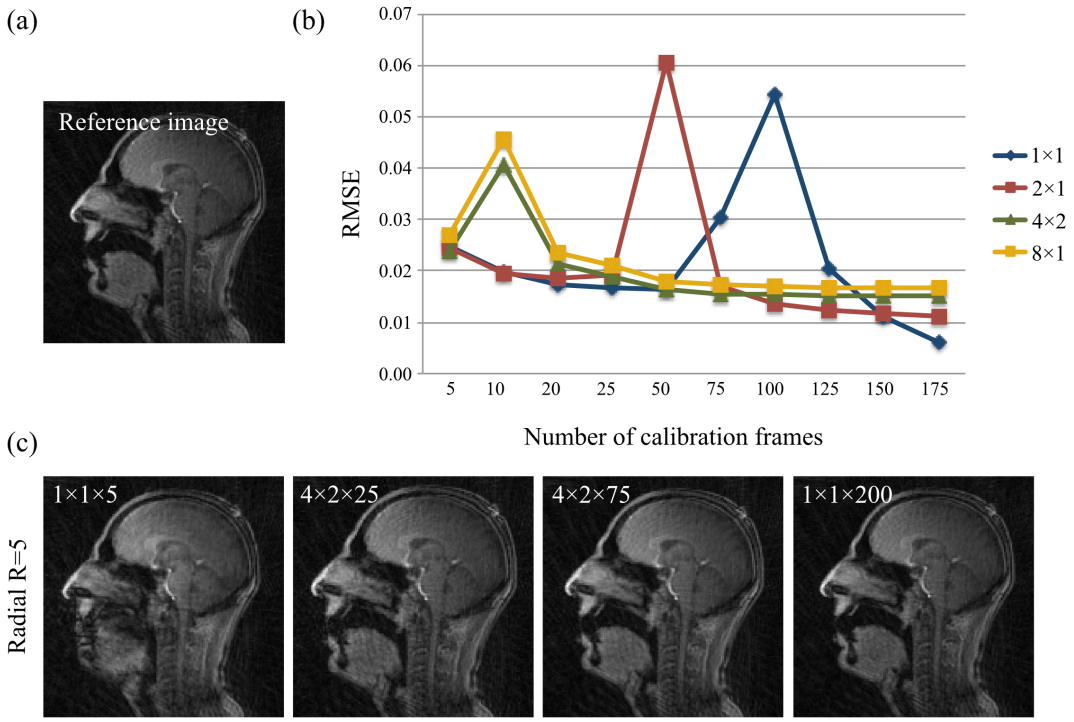


Figure 6.7: Mean RMSE from accelerated by 5-fold (R = 5 at 91 ms) radial data sets reconstructed with through-time (1×1×N) and hybrid through-time/through-k-space GRAPPA calibration methods. (a) Data used as reference reconstructed with through-time GRAPPA calibration of 1×1×200. (b) Measurements of RMSE comparing different hybrid calibration in-plane segmentation and number of calibration frames. (c) Examples of mid-sagittal dynamic frames corresponding to data reconstructed with 1×1×5, 4×2×25, 4×2×75 and 1×1×200.

accurately determined. As seen in Figure 6.8, this leads to either the incorrect reconstruction of the vocal tract anatomical shape (a) or presence of high levels of noise, if the equations are just slightly under-determined (b). However, if the number of kernel repetitions allows to accurately determine the GRAPPA weights, such as the case of 1×1×200, optimal image quality was achieved with good depiction of the velopharyngeal motion as seen by the crisp visual aspect of the intensity-time displays (d). Hybrid GRAPPA allowed maintaining a high number of kernel repetitions while reducing the need for long calibration data (e.g. 4×2×25; 25 dynamic frames instead of 200).

Measurements of intensity-time CNR of velum and neighboring air cavity undertaken in data sets reconstructed with different kernel repetitions are summarised in Table 6.2. Image data sets reconstructed with just slightly over the required minimum number of kernel repetitions ($\simeq 96$) show inferior CNR performance due to the higher random noise. Such cases correspond to GRAPPA reconstructions of 1×1×100 and 2×1×50 compared to double the kernel repetitions (200) corresponding to 1×1×200 (CNR: 4.65 ± 1.26 vs. 12.64 ± 3.60, $p<0.05$) and 2×1×100 (CNR: 4.63 ± 1.35 vs. 11.51 ± 5.32, $p<0.05$). Furthermore, due to the size of in-plane segmentation (4×2 and 8×1), reconstructions undertaken with 10 temporal frames (e.g. $4\times2\times10 = 80$) are slightly under the minimum required kernel repetitions and thus inaccurate. Within the same in-plane segmentation dimension (Table 6.2 columns), as the number of temporal frames used for the calibration increases (10 to 200), an increase in CNR performance is also noticeable. Furthermore, the highest CNR performance corresponds to the highest number of kernel occurrences (~ 1600), in this case given by 4×2×200 (16.16 ± 5.07) and 8×1×200 (15.12 ± 3.83).

Measurements of sharpness performed in the temporal and spatial direction at the velum/air boundary in data sets reconstructed with different hybrid kernel repetitions are summarised in Table 6.3. No significant difference in spatial sharpness was found between data sets reconstructed with hybrid GRAPPA cal-

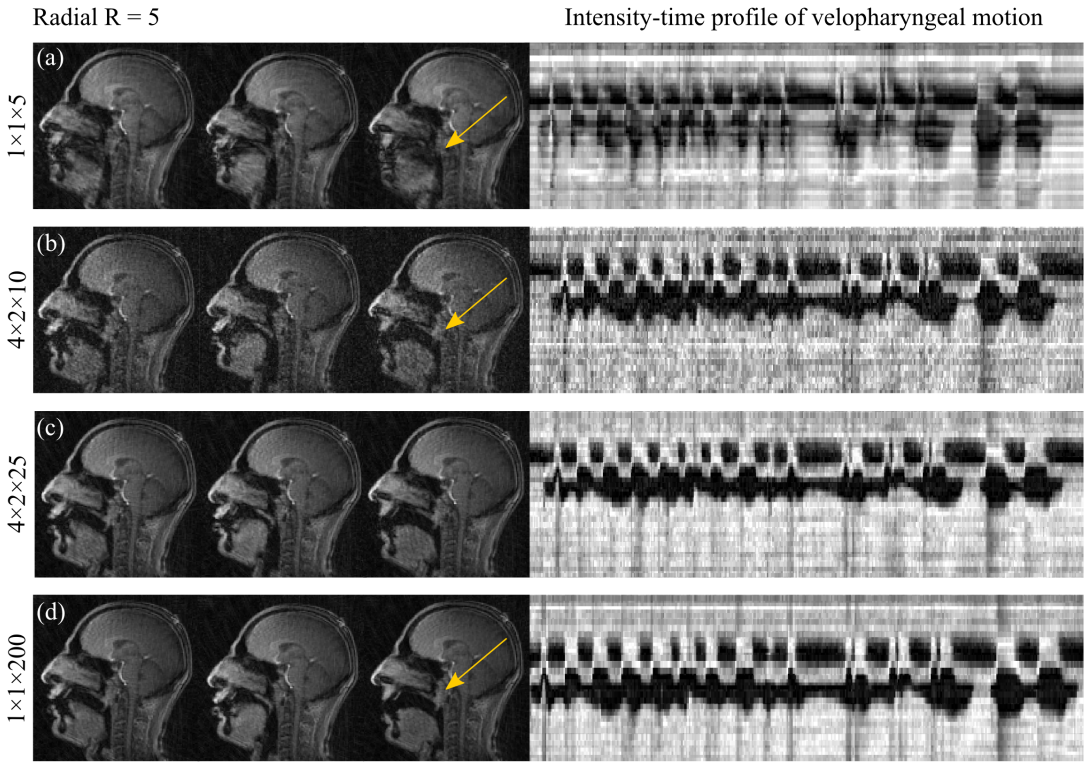


Figure 6.8: Mid-sagittal dynamic frames at 3 non-consecutive phonation moments and intensity-time profiles of velopharyngeal motion acquired with accelerated radial R = 5, reconstructed with different tt-GRAPPA calibration kernel repetition patterns. Data was reconstructed with kernel repetitions at 1×1×5, 4×2×10, 4×2×25 and 1×1×200. Weights calculations with kernel repetitions at 1×1×5 and 4×2×10 are inaccurate thus causing either major reconstruction errors (a) or increased level of noise (b). Purely through-time GRAPPA at 1×1×200 (d) shows adequate reconstruction of rt-MRI of speech and clear depiction of velopharyngeal motion. Hybrid calibration at 4×2×25 (c) provides similar image quality to the reference (1×1×200) while reducing the required number of calibration frames.

ibrations ($p = 0.78$). A pairwise comparison between data sets 1×1×100 and 1×1×200 show a significant increase in temporal sharpness from 0.05 ± 0.02 to 0.13 ± 0.09 ($p<0.05$). No other significant differences were found between pairs (column entries) in temporal sharpness.

$\mathbf{N}_{frames}$	**1×1**	**2×1**	**4×2**	**8×1**	**p-value**
10	-	-	7.61 (3.07)	6.41 (2.40)	0.59^{NS}
25	-	-	12.47 (5.01)	10.87 (2.85)	0.68^{NS}
50	-	4.63 (1.35)	13.11 (3.75)	12.81 (3.80)	0.10^{NS}
75	-	11.77 (2.42)	14.24 (3.87)	14.35 (4.20)	0.57^{NS}
100	4.65 (1.26)	11.51 (5.32)	15.75 (5.78)	14.74 (4.76)	<0.01
150	11.82 (3.72)	14.00 (5.02)	14.99 (4.44)	13.39 (7.23)	0.67^{NS}
200	12.64 (3.60)	13.47 (5.17)	16.16 (5.07)	15.12 (3.83)	0.77^{NS}
p-value	<0.1	<0.05	0.25^{NS}	0.11^{NS}	

Table 6.2: Measurements of mean CNR velum/air and standard deviation for the different calibration schemes. Measurements were performed on a short section of the speech sample, corresponding to the phonation of /zi-ni-zi/. P-values refer to ANOVA analysis of CNR measurements between the same in-plane GRAPPA segmentation (columns) or the same number of temporal frames (rows). NS-not significant.

	Temporal Sharpness	**Spatial Sharpness**
1×1×100	0.05 (0.02)^a	0.07 (0.05)
1×1×200	0.13 (0.09)	0.07 (0.02)
4×2×25	0.11 (0.06)	0.07 (0.02)
4×2×50	0.15 (0.08)	0.06 (0.02)
8×1×25	0.10 (0.06)	0.06 (0.02)
8×1×50	0.11 (0.08)	0.08 (0.03)
p-value	<0.05	0.78^{NS}

Table 6.3: Mean and standard deviation measurements of sharpness at the velum/air boundary in the time and spatial direction. Measurements were undertaken as the average of four intensity profiles drawn over the intensity-time profiles in data sets reconstructed with hybrid GRAPPA calibrations corresponding to 1×1×100, 1×1×200, 4×2×25, 4×2×50, 8×1×25 and 8×1×50. P-values correspond to ANOVA analysis undertaken comparing the temporal sharpness and spatial sharpness (columns) between different GRAPPA calibrations. Post-hoc Bonferroni paired t-test was used to identify statistically significant pairs. ^{a} $p<0.05$ pairwise comparison to 1×1×200. NS-not significant.

6.3.4 Speech tt-GRAPPA vs. commercial protocols at 3T

Speech rt-MRI acquired with commercial Cartesian SENSE and commercial radial view-sharing (Chapter 4) were compared to the proposed accelerated radial tt-GRAPPA method. Accelerated radial acquisitions (R = 5 and 10) were reconstructed offline with proposed tt-GRAPPA.

Examples of intensity-time profiles acquired at approximately 10 and 11 fps and 20 and 22.1 fps are shown in Figure 6.9.

Accelerated radial data sets reconstructed with tt-GRAPPA (1×1×200) at 11 fps demonstrated superior temporal depiction of closure events than commercial protocols, especially in closure events that required the rapid lowering and elevation of the velum (Figure 6.9 red arrows). Temporal depiction and overall visual image quality of radial tt-GRAPPA were superior to commercial radial view-sharing protocols, where closure events are blurred together and can not be adequately identified. When imaging at higher frame rates, overall quality of the commercial Cartesian intensity-time display decreases and some previously identifiable closures are now visually blended together (Figure 6.9 d). Highly accelerated radial tt-GRAPPA (R = 10, reconstructed with hybrid 2×1×200) demonstrated superior depiction of the overall closure pattern. However, due to the high acceleration factor, some aliasing is now present in the final reconstructed image.

6.4 Discussion

A through-time GRAPPA methodology to reconstruct accelerated speech real-time MRI has been demonstrated in this experimental chapter. It was shown that tt-GRAPPA could be used to greatly improve the image quality of highly accelerated real-time imaging of speech. While conventional GRAPPA relied on the repetition of the calibration kernel over k-space to acquire sufficient estimations of the GRAPPA weights, non-Cartesian tt-GRAPPA relies on the repetition of the kernel through time. Due to the intrinsic repetition of the calibration kernel

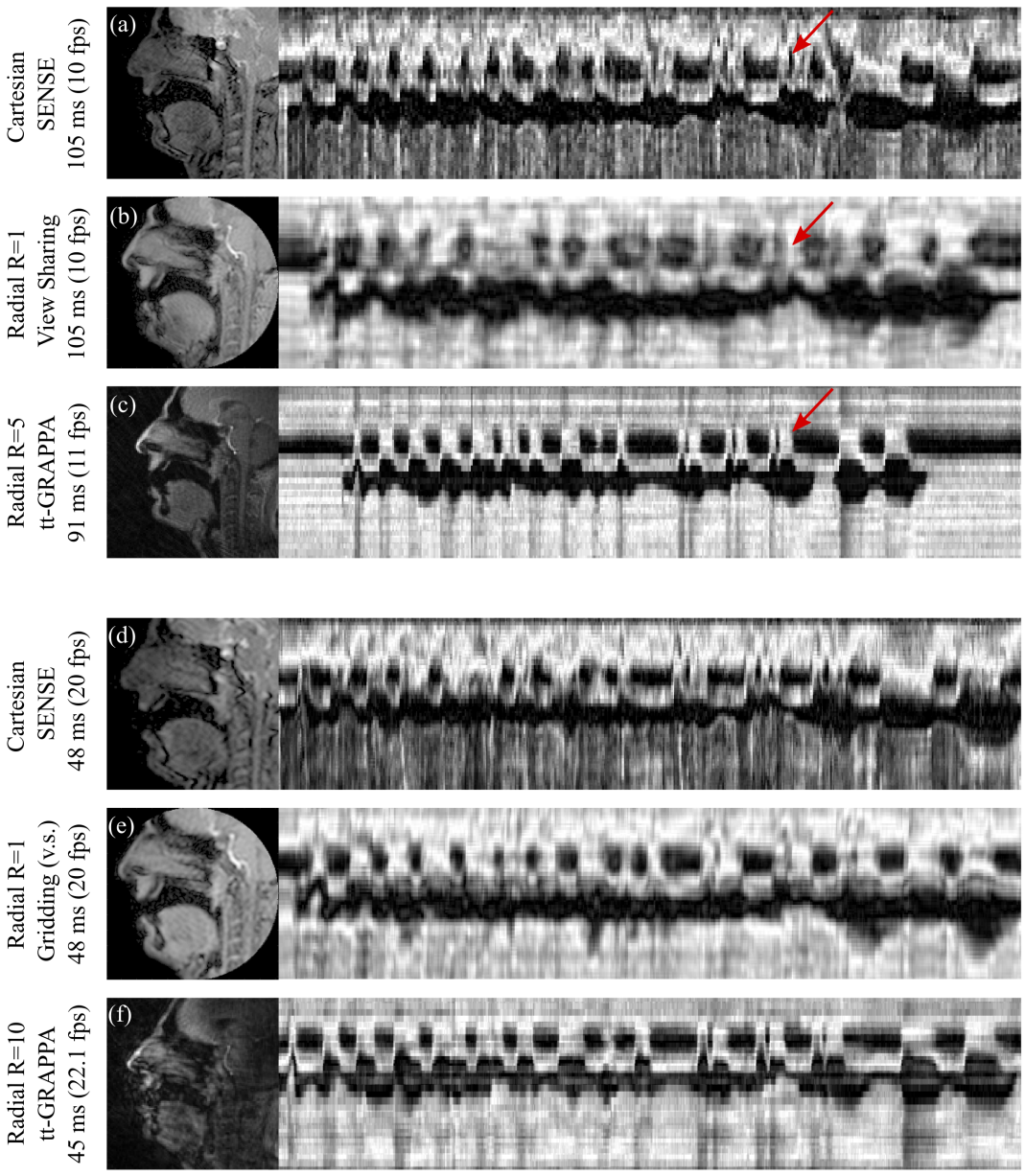


Figure 6.9: Comparison of rt-MRI of speech with (a,b) commercial Cartesian and radial view-sharing protocols previously explored at 10 fps, (c) accelerated R = 5 radial tt-GRAPPA at 11 fps, (d,e) commercial Cartesian and radial view-sharing at 20 fps and (f) accelerated R = 10 radial tt-GRAPPA at 22.1 fps. Commercial protocols were implemented as previously described in Chapter 4 and accelerated radial tt-GRAPPA were reconstructed offline with a calibration scheme 1×1×200 for R = 5 and 2×1×200 for R = 10.

through several temporal frames, tt-GRAPPA is of particular interest to improve accelerated real-time imaging of velopharyngeal motion. Temporal resolutions ranging from 4.4 to 22.1 fps at 1.9×1.9 mm^2 in-plane resolution were achieved and analysed. A preferred 5-fold acceleration, corresponding to ∼11 fps temporal resolution, was also prescribed for optimal temporal depiction of velopharyngeal motion.

Initial implementation of non-Cartesian tt-GRAPPA considered the acquisition of calibration data with up to 400 dynamic frames to perform free-breathing dynamic imaging of the heart [12]. However, this was not feasible for dynamic imaging of natural speech due to the long scanning time required for calibration (up to 1.50 minutes). Therefore, preliminary phantom experimental work was undertaken in order to investigate the effect of reducing the number of kernel repetitions through time, i.e. the number of dynamic frames acquired in the calibration data, on final image quality. Results showed a maximum RMSE, visually corresponding to a reconstructed image with high presence of random noise, when using 100 dynamic calibration frames (in-plane segmentation: 1×1). This can be explained by the fact that at least $n_c \times n_{points}$ kernel repetitions are necessary for the weights equation to be fully determined. As $n_c \times n_{points} = 16 \times (2 \times 3) = 96$, thus 100 dynamic frames is in fact the closest to the minimum kernel repetitions necessary for weight calculation. However, at this minimum point, image reconstruction is highly dependent on random noise. On the other hand, the higher the number of kernel repetitions through time (dynamic frames), the more over-determined the weight equations are, thus further improving image quality. As seen in Figure 6.3, increasing the number of calibration frames used, resulted in an overall decrease in the RMSE error. However, only negligible changes were observed past the point of about 200 dynamic frames in regards to visual image assessment and RMSE. Therefore, a value of 200 dynamic frames for the acquisition of ACS data was chosen for the final *in-vivo* validation.

Feasibility of tt-GRAPPA *in-vivo* for accelerated speech imaging was estab-

lished. Comparison of accelerated data sets at R = 5 reconstructed with zero-filled NUFFT and tt-GRAPPA demonstrated a clear improvement when using tt-GRAPPA with selected 1×1×200 reconstruction. While NUFFT data showed increased presence of image disruption due to aliasing artefacts, tt-GRAPPA provided optimal image quality with no major artefacts present and clear definition of vocal tract articulators. It also demonstrated adequate temporal depiction of velopharyngeal motion with a clear identification of all velopharyngeal closures expected from corresponding phonation sounds as shown by the intensity-time displays (Figure 6.5).

Additionally, temporal depiction of velopharyngeal closures was also investigated at different temporal resolutions, i.e. acceleration factors of radial acquisition. Results demonstrated that for slow sampling rates (R = 1, ∼2.3 fps), depiction of velopharyngeal motion was inadequate and individual closure events were not adequately identified. Although, with a 2-fold acceleration (∼4.4 fps), the overall pattern of velum motion can already be distinguished, an optimal temporal resolution was determined at 5-fold acceleration (∼11.0 fps). Velum sampling at 11 fps not only demonstrated a crisp depiction of all expected closure events, but also goes in accordance with the optimal temporal resolution recommended for velopharyngeal assessment by previous studies [41]. Lastly, high sampling rate data sets (∼22.1 fps) presented increased distortion after tt-GRAPPA reconstruction due to the high acceleration factor (R = 10). Thus, the benefits of achieving higher temporal resolutions may be outweighed by the decrease in image quality and hampered identification of velum closure events.

A major drawback of tt-GRAPPA is the high number of ACS frames required for reconstruction, thus resulting in longer scanning times. As suggested by Seiberlich et al. [12] and Lingala et al. [59], one way to overcome this is to use a so called hybrid calibration, where the weights estimation kernel is repeated not only through-time but also through-k-space (in-plane segmentation). Experimental work investigated the use of hybrid and through-time GRAPPA

calibration in final image quality and its feasibility to describe velopharyngeal motion in speech. Visual assessment of 5-fold radial data sets reconstructed with through-time GRAPPA (1×1×200) demonstrated adequate image quality for speech assessment. On the other hand, if the number of dynamic frames is insufficient to determine the GRAPPA weights correctly, image quality decreases as reconstruction errors and higher dependency on noise are introduced. Visual assessment of image data sets reconstructed with insufficient kernel repetitions (e.g. 1×1×5) showed increased reconstruction errors. In cases where the weights equation is just slightly over-determined (e.g. 1×1×100 or 2×1×50), reconstructed images showed higher dependency on random noise. On the other hand, hybrid through-k-space/through-time GRAPPA enabled maintaining a high number of kernel repetitions while reducing the need for longer calibration data (e.g. 4×2×25). Visual assessment of this data showed equivalent optimal temporal depiction of the velum with hybrid 4×2×25 reconstruction compared to reference data, 1×1×200, as shown in Figure 6.8. Additionally, no significant differences were found in temporal and spatial sharpness performance between hybrid and reference 1×1×200 GRAPPA reconstructions (Table 6.3). CNR performance was also equivalent between reference 1×1×200 (12.64 ± 3.60) and hybrid examples 2×1×100 (11.51 ± 5.32) and 4×2×50 (13.11 ± 3.75). Thus, indicating that a similar performance in CNR and sharpness of the velum can be found when using a hybrid GRAPPA compared to the through-time reference case of 1×1×200.

Lastly, the proposed accelerated radial tt-GRAPPA was compared to the commercial protocols, Cartesian and radial view-sharing, previously implemented and suggested in Chapter 4. When imaging at 10 fps, results showed that the proposed radial tt-GRAPPA scheme provided superior visual depiction of the speech closure pattern than commercial Cartesian and radial view-sharing protocols. All expected closure events were clearly identified with radial tt-GRAPPA, while this was not possible with commercial radial protocol. This was mainly due to the fact that commercial radial protocols relied on a high degree of view-sharing acceler-

ation. Although this view-sharing scheme allowed for a superior display rate, no additional data is added beyond the native frame rate (∼2.3 fps). Thus, commercial radial view-sharing protocols were intrinsically inferior as the native temporal frame rate was inadequate for velum imaging and a higher degree of view-sharing interpolation was required to achieve the final display of 10 frames per second. On the other hand, the proposed radial tt-GRAPPA presents a native frame rate of ∼11 fps. Then, missing k-space information can be recovered by means of tt-GRAPPA, resulting in a superior and temporally clearer depiction of the speech pattern. Although highly accelerated tt-GRAPPA was also undertaken, final image quality was hampered by inaccurate missing data estimation due to the high degree of under-sampling (10-fold).

Furthermore, results seem to suggest that accelerated non-Cartesian tt-GRAPPA could be ideal for future clinical translation. If a shorter ACS calibration data is acquired either pre or post standard scanning (similarly to a SENSE coil sensitivity maps reference scan in clinical Cartesian protocols), adequate image quality for velopharyngeal imaging with a hybrid calibration (e.g. 4×2×50) could be achieved. Due to the low latency reconstruction times observed in tt-GRAPPA (less than 5 minutes per complete speech task data set), on-the-fly reconstruction of accelerated data could be implemented in future work.

The main limitations of tt-GRAPPA applied to speech is the need for external calibration data that can take up to 2 minutes. However, it was also demonstrated that this could be further reduced with the use of hybrid GRAPPA. Additionally, some constraints also exist in regards to the selection of base matrix and number of radial spokes, as the number of under-sampled radial spokes must be an integer divisor of the fully-sampled number of radial spokes present in the ACS data. This requires careful planning before data can be acquired and may constrain its application to certain combinations of spatial-temporal resolution pairs.

6.5 Summary

In this chapter, the feasibility of hybrid and through-time GRAPPA in accelerated radial rt-MRI was demonstrated to assess velopharyngeal motion in a cohort of adult subjects. *In-vivo* validation of the technique prescribed a radial 5-fold acceleration protocol ($\sim$11 fps), an optimal temporal resolution according to the recommended range by previous studies [41] when assessing velopharyngeal motion. The proposed radial tt-GRAPPA scheme provided crisp temporal displays of the closure pattern where all expected closure events could be easily identified. In addition, hybrid GRAPPA allowed maintaining kernel repetitions and overall image quality (with no significant loss in CNR and sharpness performance), while reducing the number of required calibration frames and total scan time.

Chapter 7

Conclusions and future work

In this thesis, novel approaches to the imaging of the upper vocal tract during speech with real-time MRI have been presented. As the use of real-time MRI in the assessment of speech impairments is yet to be translated to clinical practice, the main focus of this experimental work was to increase clinical reliability of real-time speech MRI.

Methodologies investigated in this thesis have demonstrated that suggested rt-MRI techniques can provide adequate visual image quality and sharpness of the velum/air boundary with still a time-efficient reconstruction pipeline that could be translatable to current clinical MR systems. This might ensure that the work here presented could not only be applied in future clinical assessment of speech but also to other speech-study related areas (such as linguistic studies).

7.1 Key findings

Firstly, the optimisation and comparison of Cartesian and non-Cartesian rt-MRI protocols for the visualisation and assessment of velopharyngeal motion during speech was presented in Chapter 4. As a starting point, all imaging protocols were implemented with commercially available acquisition sequences as well as on-the-fly reconstruction methodologies provided by the scanner vendor. This not only

ensured a time-efficient and easy future translation to clinical practice as well as provided a starting point to researchers interested in the field. Imaging protocols were optimised and implemented at 1.5 T and 3 T. Quantitative and qualitative measurements such as the velum thickness in the relaxed and elevated position, dynamic CNR of the velum and surrounding air cavity and visual scoring of rt-MRI data sets were performed. Overall, spiral imaging protocols presented superior CNR velum/air performance than otherwise equivalent Cartesian protocols. In addition and particularly at 1.5 T, spiral protocols allowed to double temporal resolution (from 10 fps to 22 fps) while maintaining spatial resolution (1.9×1.9 mm^2) and with no significant loss in CNR velum/air. While quantitative measurements may provide valuable information, due to the clinical issue at focus, a qualitative visual assessment of rt-MRI protocols is essential. Analysis of visual scoring demonstrated that overall image quality at 3 T is more subject-dependent and inconsistent than at 1.5 T. While imaging at 1.5 T, a suggested spiral protocol was the recommended imaging protocol with that set-up, as it not only presented optimal image quality but also high CNR performance. On the other hand, results also suggested that when imaging at 3 T, Cartesian protocols were on average superior in visual image quality at this field strength. Due to the high inter-subject variability in image quality at 3 T, some cases may also benefit from being imaged with the spiral protocols at 3 T such as to increase CNR performance or sharpness of velum segmentation.

Therefore, in order to make recommendations on an adequate imaging protocol for velopharyngeal assessment based on the results demonstrated in Chapter 4, a two-phase approach is recommended. Firstly, a visual assessment of image quality should be considered if future clinical translation is intended. Secondly, when visual image quality is above a certain threshold and/or when more than one imaging protocol are of a similar image quality, it is advisable to take into account additional quantitative criteria such as CNR performance, improvement in sharpness of dynamic intensity-time profiles or distortions of velum thickness.

Use of such widely available methods can be a starting point to image velopharyngeal closure and future clinical translation. Additionally, the true potential of non-Cartesian imaging can be further explored with less standard reconstruction methodologies.

In the comparison of commercial protocols (Chapter 4), acceleration of non-Cartesian data is achieved with the use of a vendor provided view-sharing scheme. Although this method allows for a higher final display frame rate, no additional data is added beyond the native rate of acquisition. To fulfill the Nyquist criterion at the edges of k-space, radial sampling over-samples the centre and takes longer to fully acquire one frame than spiral counterparts. Therefore, radial protocols in the clinical set-up were less optimal, since native frame rate was intrinsically inferior (and insufficient for velopharyngeal motion sampling) and higher view-sharing interpolation was necessary to achieve equivalent temporal resolution than spiral protocols.

One possible solution to this is to under-sample k-space and reduce the total time of acquisition, that is to improve temporal resolution. Some studies have prescribed the use of iterative constrained reconstruction methodologies for the attenuation of aliasing artefacts in highly accelerated non-Cartesian acquisitions [31, 50, 63].

In Chapter 5, a first application of an iterative TGV-regularised reconstruction method to the highly accelerated imaging of velopharyngeal motion was undertaken. Selection of the TGV-regularisation weight (α) was performed by visual assessment of the reconstructed data sets due to the focus on future clinical translation. For the current study with 2.5-fold accelerated radial acquisition at 2.2×2.2 mm^2 resolution, a selected weight of 0.1 was chosen. It was demonstrated that data sets reconstructed with the chosen $\alpha = 0.1$ showed an increase in CNR performance of $\sim 111\%$ and $\sim 41\%$ respectively, compared to the non-regularised norm ($\alpha = 0$) and over-weighting of $\alpha = 1$ data sets. Overall and across the cohort, the recommended protocol at 2.5-fold radial acceleration and TGV regu-

larisation ($\alpha = 0.1$) allowed for an increase in CNR and temporal sharpness of the velum/air boundary by $\sim 47\%$ and $\sim 55\%$ respectively. Similarly to experiments undertaken in Chapter 4, increase of final display frame rate was also investigated with a view-sharing scheme. In this case, a moderate 2.5-fold acceleration (minimum for velum imaging at $\sim$7 fps) and final frame rate was increased by means of a golden-angle acquisition scheme and view-sharing reconstruction. Although data sets were acquired at $\sim$7 fps, information could be displayed at a higher temporal resolution, equivalent to highly accelerated data sets (R = 5, $\sim$12 fps), while yielding adequate image quality and diminished temporal blurring and/or artefacts due to aliasing introduced by high acceleration factors.

Although some advantages can be delivered by this methodology, required times for offline computation constitute the main limitation of this work. As explored in Chapter 5, reconstruction times for a complete speech task ran between 3 and 6 hours, depending on factor of acceleration and length of speech task. Although some decrease in computation times could be achieved with a CUDA implementation, total required times may still be prohibitive for future clinical translation where an on-the-fly reconstruction pipeline is desirable.

Therefore, the final work in this thesis focused on the investigation of a reconstruction pipeline that allowed for faster computation times (ideally closer to a on-the-fly implementation) while providing improved image quality.

A highly accelerated radial trajectory with a through-time GRAPPA reconstruction for the assessment of velopharyngeal motion was investigated. *In-vivo* results of the technique have shown an accelerated radial protocol ($\sim$11 fps), an optimal temporal resolution when assessing velopharyngeal motion [41]. The proposed radial tt-GRAPPA scheme (through-time calibration at $1\times1\times200$) provided crisp intensity-time displays of the closure pattern, where all expected closure events could be easily identified.

One major disadvantage of through-time GRAPPA is the high number of calibration frames required, thus increasing total scanning times. Thus, a hybrid

GRAPPA (with through-k-space and through-time kernel repetition) was investigated. This allowed maintaining a high number of kernel repetitions while reducing the need for longer calibration data. Data sets reconstructed with hybrid 2×1×100 and 4×2×25 demonstrated equivalent CNR performance than original through-time calibration 1×1×200. Additionally, the proposed radial tt-GRAPPA scheme provided superior visual depiction of the speech closure pattern than commercial Cartesian and radial view-sharing protocols (Chapter 4). Furthermore, due to the low latency reconstruction times observed in tt-GRAPPA, on-the-fly reconstruction of accelerated data could be implemented in future work.

In conclusion, the experimental work presented in this thesis made novel contributions to the field of real-time MRI of speech. **(i)** Optimised imaging protocols have been presented for the assessment of velopharyngeal motion at 1.5 T and 3 T with commercially available resources. **(ii)** Further improvements in image quality and temporal depiction of velopharyngeal motion were then described with the use of accelerated non-Cartesian imaging and iterative TGV regularised reconstruction methodology. However, due to the high computation times required, this approach was deemed inadequate for future clinical translation and a second approach was investigated. **(iii)** An accelerated radial through-time GRAPPA was then suggested with great potential for clinical translation. The proposed radial tt-GRAPPA approach presented in Chapter 6 provided not only adequate image quality and temporal depiction of the velum motion as well as fast computation times that could make it easily translatable to an on-the-fly reconstruction pipeline.

7.2 Future work

Some future work directions can be suggested following the work here presented and current gaps observed in literature of the field.

Due to the low latency reconstruction times demonstrated by the tt-GRAPPA

reconstruction approach, an on-the-fly reconstruction pipeline could be implemented in the future. Open-source tools such as GPI and BART (Berkeley Advanced Reconstruction Toolbox) [82, 83] and Gadgetron [84] could for example serve as platform to implement the on-the-fly reconstruction pipeline. This would allow to improve image quality and temporal depiction of the velopharyngeal pattern compared to conventional commercial protocols while still aiming for an on-the-fly scanning and speech assessment with a speech and language therapist present in the scanner room. Additionally, the use of tt-GRAPPA is not limited to the use of a radial acquisition trajectory and could be further extended to other non-Cartesian imaging protocols (e.g. spirals). This could allow for further temporal acceleration, particularly required when performing faster speech tasks or imaging other vocal tract articulators (e.g. rapid tongue movements) [59].

Furthermore, improvements in temporal acceleration could also be used to increase the number of imaging planes as suggested in Section 3.4.4. A main advantage of MRI compared to conventional x-ray videofluoroscopic examinations is the ability to acquired multiple image planes without repositioning the patient or limitation due to radiation exposure. Multi-slice acquisitions have previously been implemented in speech research settings and proven to provide valuable information (e.g. lateral motion of the pharynx) that would not be visible in the mid-sagittal plane only [61].

Therefore, preliminary experimental work undertaken after the completion of this thesis is currently investigating the use of the proposed radial tt-GRAPPA scheme on an interleaved 2-slice imaging protocol (Figure 7.1). This would allow it to image real-time velopharyngeal motion from two simultaneous imaging planes, a mid-sagittal and an oblique plane.

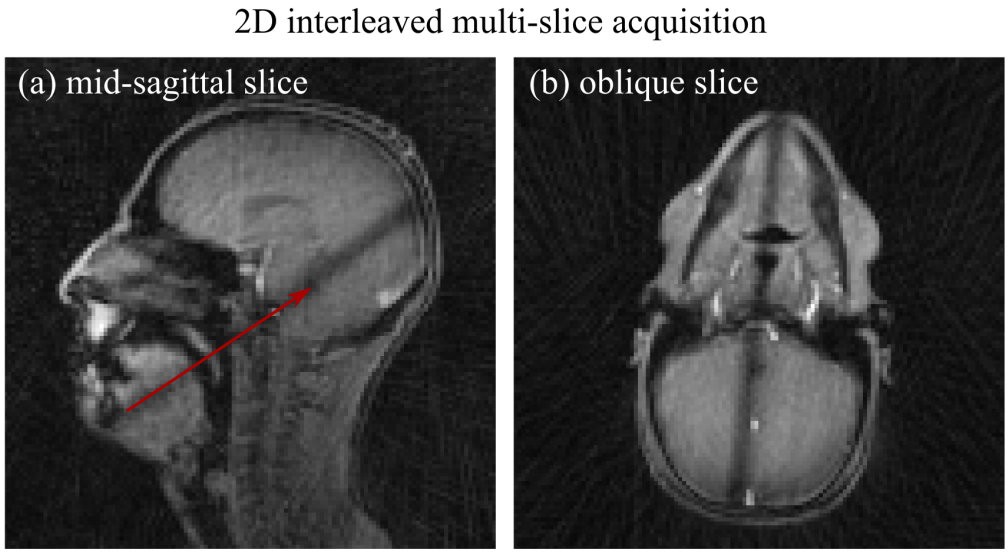


Figure 7.1: Preliminary results from a 2D interleaved multi-slice acquisition with accelerated radial tt-GRAPPA. This allows for an interleaved imaging of the velopharyngeal port in two imaging planes, (a) a mid-sagittal and (b) an oblique plane. Scan planning of the oblique plane is performed along the main direction of velum motion in the mid-sagittal plane (red arrow).

Supplementary Media Files

Supplementary media files can be found attached to the digital format of this thesis. These refer to the experimental work undertaken in this thesis as follows:

S1 - Example of commercial Cartesian, radial and spiral real-time speech MRI at 10 fps at 3 T.

S2 - Example of commercial Cartesian, radial and spiral real-time speech MRI at 20 fps at 3 T.

S3 - Example of accelerated radial rt-MRI of speech reconstructed with TGV-regularised iterative method.

S4 - Example of highly accelerated radial rt-MRI of speech reconstructed with offline through-time GRAPPA.

S5 - Comparison of image quality between commercial Cartesian SENSE, commercial radial view-sharing and proposed accelerated radial tt-GRAPPA.

List of Abbreviations

ANOVA - Analysis Of VAriance

CNR - Contrast to Noise Ratio

FFE - Fast Field Echo

FID - Free induction decay

FLASH - Fast Low Angle Shot

FOV - Field-Of-View

GA - Golden-Angle

GRAPPA - Generalized Autocalibrating Partially Parallel Acquisition

GRE - Gradient Echo

MRI - Magnetic Resonance Imaging

RF - Radio-frequency

RMSE - Root mean squared error

rt-MRI - real-time Magnetic Resonance Imaging

SE - Spin Echo

SENSE - Sensitivity Encoding

SNR - Signal-to-noise ratio

SSFP - Steady State Free Precession

TE- Echo Time

TGV - Total Generalized Variation

TR - Repetition Time

tt-GRAPPA - Through-time GRAPPA

TV - Total Variation

VPI - Velopharyngeal Insufficiency

Bibliography

[1] Ipdtoc Working Group. Prevalence at birth of cleft lip with or without cleft palate: data from the International Perinatal Database of Typical Oral Clefts (IPDTOC). *The Cleft Palate-Craniofacial Journal*, 48(1):66–81, 2011.

[2] Kim Bettens, Floris L Wuyts, and Kristiane M Van Lierde. Instrumental assessment of velopharyngeal function and resonance: A review. *Journal of Communication Disorders*, 2014.

[3] Andrew D Scott, Marzena Wylezinska, Malcolm J Birch, and Marc E Miquel. Speech MRI: morphology and function. *Physica Medica*, 30(6):604–618, 2014.

[4] Andrew D Scott, R Boubertakh, Malcom J Birch, and Marc E Miquel. Towards clinical assessment of velopharyngeal closure using MRI: evaluation of real-time MRI sequences at 1.5T and 3T. *The British Journal of Radiology*, 85(1019):1083–1092, 2012.

[5] Tobias Breyer, Matthias Echternach, Susan Arndt, Bernhard Richter, Oliver Speck, Martin Schumacher, and Michael Markl. Dynamic magnetic resonance imaging of swallowing and laryngeal motion using parallel imaging at 3 T. *Magnetic Resonance Imaging*, 27(1):48–54, 2009.

[6] Youkyung Bae, David P Kuehn, Charles A Conway, and Bradley P Sutton. Real-time magnetic resonance imaging of velopharyngeal activities with simultaneous speech recordings. *The Cleft Palate-Craniofacial Journal*, 48(6):695–707, 2011.

[7] Shrikanth Narayanan, Krishna Nayak, Sungbok Lee, Abhinav Sethy, and Dani Byrd. An approach to real-time magnetic resonance imaging for speech production. *Journal of Acoustical Society of America*, 115(4):1171–1776, 2004.

[8] Aaron Niebergall, Shuo Zhang, Esther Kunay, Gotz Keydana, Michael Job, Martin Uecker, and Jens Frahm. Real time MRI of speaking at a resolution of 33 ms: undersampled radial FLASH with nonlinear inverse reconstruction. *Magnetic Resonance in Medicine*, 69(2):477–485, 2013.

[9] Ambros J Beer, Paul Hellerhoff, Angela Zimmermann, Katalin Mady, Robert Sader, Ernst J Rummeny, and Christian Hannig. Dynamic near-real-time magnetic resonance imaging for analyzing the velopharyngeal closure in comparison with videofluoroscopy. *Journal of Magnetic Resonance Imaging*, 20(5):791–797, 2004.

[10] Pallavi Sagar and Katherine Nimkin. Feasibility study to assess clinical applications of 3-T cine MRI coupled with synchronous audio recording during speech in evaluation of velopharyngeal insufficiency in children. *Pediatric Radiology*, 45(2):217–227, 2014.

[11] Amanda L Silver, Katherine Nimkin, Jean E Ashland, Satrajit S Ghosh, Andre J W van der Kouwe, Matthew T Brigger, and Christopher J Hartnick. Cine magnetic resonance imaging with simultaneous audio to evaluate pediatric velopharyngeal insufficiency. *Archives of otolaryngology–head & neck surgery*, 137(3):258–263, 2011.

[12] Nicole Seiberlich, Philipp Ehses, Jeff Duerk, Robert Gilkeson, and Mark Griswold. Improved radial GRAPPA calibration for real-time free-breathing cardiac imaging. *Magnetic Resonance in Medicine*, 65(2):492–505, 2011.

[13] Donald B Plewes and Walter Kucharczyk. Physics of MRI: a primer. *Journal of Magnetic Resonance Imaging*, 35(5):1038–54, 2012.

[14] Anthony A Mancuso. *Magnetic Resonance Imaging: Physical Principles and Applications*, volume 99. Academic Press Inc., 1992.

[15] Stig Ljunggren. A simple graphical representation of Fourier-based imaging methods. *Journal of Magnetic Resonance Imaging*, 54:338–343, 1983.

[16] Matt A Bernstein, Kevin King, and Xiaohong Joe Zhou. *Handbook of MRI Pulse Sequences*. Elsevier Inc., 2004.

[17] HY Carr. Steady-state free precession in nuclear magnetic resonance. *Phys Rev*, 112:1963–1701, 1958.

[18] Klaus Scheffler and Stefan Lehnhardt. Principles and applications of balanced SSFP techniques. *European Radiology*, 13(11):2409–18, 2003.

[19] C B Ahn, J H Kim, and Z H Cho. High-speed spiral-scan echo planar NMR imaging-I. *IEEE transactions on medical imaging*, 5(1):2–7, 1986.

[20] J A OSullivan. A fast sinc function gridding algorithm for Fourier inversion in computer tomography. *IEEE transactions on medical imaging*, 4(4):200–207, 1985.

[21] Kenneth O Johnson and James G Pipe. Convolution kernel design and efficient algorithm for sampling density correction. *Magnetic Resonance in Medicine*, 61(2):439–47, 2009.

[22] John I Jackson, Craig H Meyer, Dwight G Nishimura, and Albert Macovski. Selection of a Convolution Function for Fourier Inversion Using Gridding. *IEEE transactions on medical imaging*, 1(3):200–207, 1991.

[23] Philip J Beatty, Dwight G Nishimura, and John M Pauly. Rapid gridding reconstruction with a minimal oversampling ratio. *IEEE transactions on medical imaging*, 24(6):799–808, 2005.

[24] Jeffrey A. Fessler and Bradley P. Sutton. Nonuniform Fast Fourier Transforms Using Min-Max Interpolation. *IEEE Transactions on Signal Processing*, 51(2):560–574, 2003.

[25] Thomas F Budinger, Hubertus Fischer, Dietmar Hentschel, Hans-Erich Reinfelder, and Franz Schmitt. Physiological effects of fast oscillating magnetic field gradients. *Journal of Computer Assisted Tomography*, 15(6):909–914, 1991.

[26] Katherine L Wright, Jesse I Hamilton, Mark A Griswold, Vikas Gulani, and Nicole Seiberlich. Non-Cartesian parallel imaging reconstruction. *Journal of Magnetic Resonance Imaging*, 40:1022–1040, 2014.

[27] David J Larkman and Rita G Nunes. Parallel magnetic resonance imaging. *Physics in Medicine and Biology*, 52(7):R15–55, 2007.

[28] Anagha Deshmane, Vikas Gulani, Mark A Griswold, and Nicole Seiberlich. Parallel MR Imaging. *Journal of Magnetic Resonance Imaging*, 36:55–72, 2012.

[29] K P Pruessmann, M Weiger, M B Scheidegger, and P Boesiger. SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–62, 1999.

[30] Mark A Griswold, Peter M Jakob, Robin M Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–10, 2002.

[31] Michael Burdumy, Louisa Traser, Bernhard Richter, Matthias Echternach, Jan G Korvink, Jürgen Hennig, and Maxim Zaitsev. Acceleration of MRI of the vocal tract provides additional insight into articulator modifications. *Journal of Magnetic Resonance Imaging*, 42(4):925–35, 2015.

[32] Patrick Lynch. Head mid sagittal view with velopharyngeal space. *Wikipedia Commons Illustrations*, 2006.

[33] Peter A Mossey, Julian Little, Ron G Munger, Mike J Dixon, and William C Shaw. Cleft lip and palate. *Lancet*, 374(9703):1773–85, 2009.

[34] Debbie Sell and Valerie Pereira. Instrumentation in the analysis of the structure and function of the velopharyngeal mechanism. In Sarah Howard and Anette Lohmander, editors, *Cleft Palate Speech: assessment and intervention*, pages 145–162. Wiley-Blackwell, 2012.

[35] Christina Havstam, Anette Lohmander, Christina Persson, Hans Dotevall, Agneta Lith, and Jan Lilja. Evaluation of VPI-assessment with videofluoroscopy and nasoendoscopy. *British Journal of Plastic Surgery*, 58(7):922–31, 2005.

[36] Kelly A Weir, Sandra M McMahon, Gillian Long, Judith A Bunch, Nirmala Pandeya, Kerry S Coakley, and Anne B Chang. Radiation doses to children during modified barium swallow studies. *Pediatric Radiology*, 37(3):283–90, 2007.

[37] M.T. Crawley, P. Savage, and F. Oakley. Patient and operator dose during fluoroscopic examination of swallow mechanism. *British Journal of Radiology*, (77):654–656, 2004.

[38] Karen A. Golding-Kushner. Standardization for the reporting of nasopharyngoscopy and multiview videofluoroscopy: a report from an international working group. *Cleft Palate Journal*, 27(4):337–348, 1990.

[39] Sajan Goud Lingala, Brad P Sutton, Marc E Miquel, and Krishna S Nayak. Recommendations for real-time speech MRI. *Journal of Magnetic Resonance Imaging*, 43(1):28–44, 2016.

[40] David P. Kuehn. A cineradiographic investigation of velar movement variables in two normals. *Cleft Palate Journal*, 13:88–103, 1976.

[41] Matthieu Ruthven, Andreia C. Freitas, Stephen F. Keevil, and Marc E. Miquel. Real-time speech MRI: what is the optimal temporal resolution for clinical velopharyngeal closure assessment? In *Proc. Intl. Soc. Mag. Reson. Med.*, page 206, 2016.

[42] Paula Martins, Inês Carbone, Alda Pinto, Augusto Silva, and António Teixeira. European Portuguese MRI based speech production studies. *Speech Communication*, 50(11-12):925–952, 2008.

[43] Matthias Echternach, Johan Sundberg, Susan Arndt, Michael Markl, Martin Schumacher, and Bernhard Richter. Vocal tract in female registers- A dynamic real-time MRI study. *Journal of Voice*, 24(2):133–139, 2010.

[44] Matthias Echternach, Johan Sundberg, Susan Arndt, Tobias Breyer, Michael Markl, Martin Schumacher, and Bernhard Richter. Vocal tract and register changes analysed by real-time MRI in male professional singers-a pilot study. *Logopedics, phoniatrics, vocology*, 33(2):67–73, 2008.

[45] Christiane Kulinna-Cosentini, Christian Czerny, Arnulf Baumann, Michael Weber, and Klaus Sinko. TrueFisp versus HASTE sequences in 3T cine MRI: Evaluation of image quality during phonation in patients with velopharyngeal insufficiency. *European Radiology*, 10.1007/s0, 2015.

[46] Kate McLeish, Sebastian Kozerke, William R Crum, and Derek L G Hill. Free-breathing radial acquisitions of the heart. *Magnetic Resonance in Medicine*, 52(5):1127–35, 2004.

[47] Angela O Leung, Ian Paterson, and Richard B Thompson. Free-breathing cine MRI. *Magnetic Resonance in Medicine*, 60(3):709–717, 2008.

[48] Shuo Zhang, Kai Tobias Block, and Jens Frahm. Magnetic resonance imaging in real time: advances using radial FLASH. *Journal of Magnetic Resonance Imaging*, 31(1):101–109, 2010.

[49] Martin Uecker, Thorsten Hohage, Kai Tobias Block, and Jens Frahm. Image reconstruction by regularized nonlinear inversion–joint estimation of coil sensitivities and image content. *Magnetic Resonance in Medicine*, 60(3):674–82, 2008.

[50] Martin Uecker, Shuo Zhang, and Jens Frahm. Nonlinear inverse reconstruction for real-time MRI of the human heart using undersampled radial FLASH. *Magnetic Resonance in Medicine*, 63(6):1456–62, 2010.

[51] Joachim Lotz, Jan M Sohns, Michael Steinmetz, Johannes T Kowallick, Christina Schulte, Wieland Staab, Arun Joseph, Klaus-Dietmar Merboldt, Dirk Voit, Shuo Zhang, Martin Uecker, Christina Unterberg-Buchwald, Gerd Hasenfus, and Jens Frahm. High resolution real-time CMR of function and flow: initial clinical results. *Journal of Cardiovascular Magnetic Resonance*, 15(Suppl 1):E99, 2013.

[52] Martin Uecker, Shuo Zhang, Dirk Voit, Klaus-dietmar Merboldt, and Jens Frahm. Real-time MRI: recent advances using radial FLASH. *Imaging in Medicine*, 4(4):461–476, 2012.

[53] KS Nayak, JM Pauly, PC Yang, BS Hu, CH Meyer, and DG Nishimura. Real-Time Interactive Coronary MRA. *Magnetic Resonance in Medicine*, 46:430–435, 2001.

[54] Krishna S Nayak, Charles H Cunningham, Juan M Santos, and John M Pauly. Real-time cardiac MRI at 3 tesla. *Magnetic Resonance in Medicine*, 51(4):655–60, 2004.

[55] Vikram Ramanarayanan, Loius Goldstein, Dani Byrd, and Shrikanth Narayanan. An investigation of articulatory setting using real-time magnetic resonance imaging. *Journal of Acoustical Society of America*, 134(1):510–519, 2013.

[56] Michael Proctor, Erik Bresch, Dani Byrd, Krishna Nayak, and Shrikanth Narayanan. Paralinguistic mechanisms of production in human beatboxing : A real-time magnetic resonance imaging study. *Journal of Acoustical Society of America*, 133(2):1043–1054, 2013.

[57] Bradley P Sutton, Jeffrey Tsao, H Shinagawa, and David P Kuehn. Dynamic imaging of muscles during speech using interleaved spiral FLASH. In *Proc. Intl. Soc. Mag. Reson. Med.*, page 3377, 2006.

[58] Nicole Seiberlich, Gregory Lee, Philipp Ehses, Jeffrey L Duerk, Robert Gilkeson, and Mark A Griswold. Improved temporal resolution in cardiac imaging using through-time spiral GRAPPA. *Magnetic Resonance in Medicine*, 66(6):1682–8, 2011.

[59] Sajan Goud Lingala, Yinghua Zhu, Yongwan Lim, Asterios Toutios, Yunhua Ji, Wei-Ching Lo, Nicole Seiberlich, Shrikanth Narayanan, and Krishna S Nayak. Feasibility of through-time spiral generalized autocalibrating partial parallel acquisition for low latency accelerated real-time MRI of speech. *Magnetic Resonance in Medicine*, 78(6):2275–2282, 2017.

[60] C. Drissi, M. Mitrofanoff, C. Talandier, C. Falip, V. Le Couls, and C. Adamsbaum. Feasibility of dynamic MRI for evaluating velopharyngeal insufficiency in children. *European Radiology*, 21(7):1462–1469, 2011.

[61] Yoon-Chul Kim, Michael I Proctor, Shrikanth S Narayanan, and Krishna S Nayak. Improved imaging of lingual articulation using real-time multislice MRI. *Journal of Magnetic Resonance Imaging*, 35(4):943–8, 2012.

[62] YC Kim, MI Proctor, SS Narayanan, and KS Nayak. Visualization of Vocal Tract Shape Using Interleaved Real-Time MRI of Multiple Scan Planes. In *InterSpeech*, pages 269–272, 2011.

[63] Sajan Goud Lingala, Yinghua Zhu, Yoon-Chul Kim, Asterios Toutios, Shrikanth Narayanan, and Krishna S Nayak. A fast and flexible MRI system for the study of dynamic vocal tract shaping. *Magnetic Resonance in Medicine*, 77(1):112–125, 2016.

[64] Maojing Fu, Bo Zhao, Christopher Carignan, Ryan K. Shosted, Jamie L. Perry, David P. Kuehn, Zhi-Pei Liang, and Bradley P. Sutton. High-resolution dynamic speech imaging with joint low-rank and sparsity constraints. *Magnetic Resonance in Medicine*, 73(5):1820–1832, 2015.

[65] Marc E Miquel, Marzena Wylezinska-Arridge, M. Pinkstone, C. Theobald, Malcolm J Birch, and Andrew D Scott. Assessment of velopharyngeal closure and soft palate anatomy using MRI in cleft palate patients. *Med Phys Int*, 1:459, 2013.

[66] SL Ettema, DP Kuehn, AL Perlman, and N Alperin. Magnetic resonance imaging of the levator veli palatini muscle during speech. *The Cleft Palate-Craniofacial Journal*, 39(2):130–144, 2002.

[67] Jamie L Perry. Variations in velopharyngeal structures between upright and supine positions using upright magnetic resonance imaging. *The Cleft Palate-Craniofacial Journal*, 48(2):123–33, 2011.

[68] MR Hestenes and E Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–436, 1952.

[69] KT Pruessmann, M Weiger, P Bornet, and P Boesiger. Advances in Sensitivity Encoding With Arbitrary k -Space Trajectories. *Magnetic Resonance in Medicine*, 651(2):638–651, 2001.

[70] Kai Tobias Block, Martin Uecker, and Jens Frahm. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine*, 57(6):1086–98, 2007.

[71] Kai Tobias Block. *Advanced Methods for Radial Data Sampling in Magnetic Resonance Imaging*. PhD thesis, 2008.

[72] AN Tikhonov and VI Arsenin. Solutions of ill-posed problems. 1977.

[73] LI Rudin, S Osher, and E Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259 – 262, 1992.

[74] Florian Knoll, Kristian Bredies, Thomas Pock, and Rudolf Stollberger. Second order total generalized variation (TGV) for MRI. *Magnetic Resonance in Medicine*, 65(2):480–91, 2011.

[75] Michael Burdumy. *Dynamic Imaging of Singers Using Magnetic Resonance Tomography*. PhD thesis, 2016.

[76] Kristian Bredies, Karl Kunisch, and Thomas Pock. Total Generalized Variation. *SIAM Journal of Imaging Sciences*, 3:492–526, 2010.

[77] Robert Grimm. *Reconstruction Techniques for Dynamic Radial MRI*. PhD thesis, 2015.

[78] Yoon Chul Kim, Shrikanth S Narayanan, and Krishna S Nayak. Flexible retrospective selection of temporal resolution in real-time speech MRI using a golden-ratio spiral view order. *Magnetic Resonance in Medicine*, 65(5):1365–1371, 2011.

[79] Katherine L Wright, Gregory R Lee, Philipp Ehses, Mark a Griswold, Vikas Gulani, and Nicole Seiberlich. Three-dimensional through-time radial GRAPPA for renal MR angiography. *Journal of Magnetic Resonance Imaging*, 40(4):864–874, 2014.

[80] Yong Chen, Gregory R Lee, Gunhild Aandal, Chaitra Badve, Katherine L Wright, Mark A Griswold, Nicole Seiberlich, and Vikas Gulani. Rapid Volumetric T1 Mapping of the Abdomen Using Three-Dimensional Through-Time Spiral GRAPPA. *Magnetic Resonance in Medicine*, 1465:1457–1465, 2016.

[81] Lingala Sajan, Yinghua Zhu, Yunhua Ji, Asterios Toutios, Wei-Ching Lo, Nicole Seiberlich, Shrikanth S. Narayanan, and Krishna S. Nayak. Accelerating Real-time MRI of speech using spiral through-time GRAPPA. In *Proc. Intl. Soc. Mag. Reson. Med.*, pages 8–10, 2016.

[82] Jonathan I Tamir, Frank Ong, Joseph Y Cheng, and Michael Lusting. Generalized Magnetic Resonance Image Reconstruction using The Berkeley Advanced Reconstruction Toolbox. *ISMRM Workshop on Data Sampling and Image Reconstruction*, 1, 2015.

[83] Martin Uecker, Frank Ong, Jonathan I Tamir, Dara Bahri, Patrick Virtue, Joseph Y Cheng, Tao Zhang, and Michael Lustig. Berkeley Advanced Reconstruction Toolbox. *Proc. Intl. Soc. Mag. Reson. Med.*, 23:2486, 2015.

[84] Michael S Hansen and Thomas S Sorensen. Gadgetron: An Open Source Framework for Medical Image Reconstruction. *Magnetic Resonance in Medicine*, 69:1768–1776, 2013.